Facial Recognition Vendor Test 2000

Evaluation Report

February 16, 2001

Duane M. Blackburn
NAVSEA Dahlgren Division
17320 Dahlgren Road
Dahlgren, VA 22448
blackburndm@nswc.navy.mil

Mike Bone
NAVSEA Crane Division
300 Highway 361
Crane, IN 47522
bone_mike@crane.navy.mil

P. Jonathon Phillips, Ph.D.
Defense Advanced Research Projects Agency
3701 N. Fairfax Drive
Arlington, VA 22203
jphillips@darpa.mil

Sponsored by:
DoD Counterdrug Technology Development Program Office
Defense Advanced Research Projects Agency
National Institute of Justice

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
16022001	Unknown	-

Title and Subtitle Face Recognition Vendor Test 2000: Evaluation Report	**Contract Number**
	Grant Number
	Program Element Number
Author(s) Duane M. Blackburn, P. Jonathon Phillips, Mike Bone	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) DARPA 3701 North Fairfax Dr. Arlington, VA 22203 National Institute of Justice DoD Counterdrug Technology Development Program Office 17320 Dahlgren Rd, Code B07 Dahlgren, VA 22448	**Performing Organization Report Number**
Sponsoring/Monitoring Agency Name(s) and Address(es)	**Sponsor/Monitor's Acronym(s)**
	Sponsor/Monitor's Report Number(s)
Distribution/Availability Statement Approved for public release, distribution unlimited	
Supplementary Notes The original document contains color images.	

Abstract

The biggest change in the facial recognition community since the completion of the FERET program has been the introduction of facial recognition products to the commercial market. Open market competitiveness has driven numerous technological advances in automated face recognition since the FERET program and signifi cantly lowered system costs. Today there are dozens of facial recognition systems available that have the potential to meet performance requirements for numerous applications. But which of these systems best meet the performance requirements for given applications? Repeated inquiries from numerous government agencies on the current state of facial recognition technology prompted the DoD Counterdrug Technology Development Program Offi ce to establish a new set of evaluations. The Facial Recognition Vendor Test 2000 (FRVT 2000) was cosponsored by the DoD Counterdrug Technology Development Program Offi ce, the National Institute of Justice and the Defense Advanced Research Projects Agency and was administered in May and June 2000.

Subject Terms

Report Classification unclassified	**Classification of this page** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** SAR
Number of Pages 70	

Acknowledgements

The sponsors of the Facial Recognition Vendor Test 2000 would like to thank the following individuals for their assistance throughout these evaluations.

- Mr. Patrick Grother of the National Institute of Standards and Technology for providing the software tools and consultation that enabled the authors to perform this evaluation.

- Mr. Tom Coty and Mr. Chris Miles from the National Institute of Justice, and Mr. George Lukes from the Defense Advanced Research Projects Agency for their support throughout the FRVT 2000 evaluations and for reviewing this document.

- Mrs. Jo Gann, Office of the National Drug Control Policy for her advice in general and for reviewing this document.

- Mrs. Kim Shepard from BRTRC for providing the evaluation web sites and the many changes that followed.

- Ms. Elaine Newton with RAND Corp. for reviewing this document.

- Ms. Kathy Cole of Schafer Corp. for preparing the final document.

Executive Overview

1 Introduction

The biggest change in the facial recognition community since the completion of the FERET program has been the introduction of facial recognition products to the commercial market. Open market competitiveness has driven numerous technological advances in automated face recognition since the FERET program and significantly lowered system costs. Today there are dozens of facial recognition systems available that have the potential to meet performance requirements for numerous applications. But which of these systems best meet the performance requirements for given applications?

Repeated inquiries from numerous government agencies on the current state of facial recognition technology prompted the DoD Counterdrug Technology Development Program Office to establish a new set of evaluations. The Facial Recognition Vendor Test 2000 (FRVT 2000) was cosponsored by the DoD Counterdrug Technology Development Program Office, the National Institute of Justice and the Defense Advanced Research Projects Agency and was administered in May and June 2000.

2 Goals of the FRVT 2000

The sponsors of the FRVT 2000 had two major goals for the evaluation. The first was a technical assessment of the capabilities of commercially available facial recognition systems. They wanted to know the strengths and weaknesses of each individual system and obtain an understanding of the current state of the art for facial recognition.

The second goal was to educate the biometrics community and the general public on how to present and analyze results. The sponsors had seen vendors and would-be customers quote outstanding performance specifications without understanding that these specifications are virtually useless without knowing the details of the test that was used to produce the quoted results.

3 FRVT 2000 Evaluation Methodology

The FRVT 2000 was based on the evaluation methodology proposed in "An Introduction to Evaluating Biometric Systems," by P. J. Phillips, A. Martin, C. L. Wilson and M. Przybocki in *IEEE Computer*, February 2000, pp. 56-63. This methodology proposes a three-step evaluation protocol: a top-level technology evaluation, followed by a scenario evaluation and an operational evaluation.

3.1 Recognition Performance Test (A Technology Evaluation)

The goal of a technology evaluation is to compare competing algorithms from a single technology, which in this case is facial recognition. Testing of all algorithms is done on a standardized database collected by a universal sensor and should be performed by an organization that will not see any benefit should one algorithm outperform the others. The use of a test set ensures that all participants see the same data. Someone with a need for facial recognition can look at the results from the images that most closely resemble their situation and can determine, to a reasonable extent, what results they should expect.

The operation of the Recognition Performance Test in the FRVT 2000 was very similar to the original FERET evaluations that were sponsored by the DoD Counterdrug Technology Development Program Office. Vendors were given 13,872 images and were asked to compare each image to all of the other images (more than 192 million comparisons). This data was used to form experiments that will

show how well the systems respond to numerous variables such as pose, lighting, and image compression level.

3.2 Product Usability Test (A Limited Example of a Scenario Evaluation)

A scenario evaluation is an evaluation of the complete facial recognition system, rather than the facial recognition algorithm only. The participating vendors were allowed to choose the components (such as camera, lighting and the like) that they would normally recommend for this scenario. These components play a major role in the ability of a facial recognition system to successfully operate in a live environment. Therefore, it was imperative that these components, and their interactions, be evaluated as a system using live test subjects.

The Product Usability Test is an example of a limited scenario evaluation. A full scenario evaluation would have used significantly more test subjects and lasted a period of weeks, but it would have also been done on only one or two systems. The participating vendors were not paid to have their systems evaluated for the FRVT 2000 so it would have been unfair to ask each of them to spend their own money to support a multiweek evaluation. The scenario chosen for the FRVT 2000 Product Usability Test was access control.

The Product Usability Tests consisted of two timed test, which were used to measure the response time of the overall system for two operational scenario simulations: the Old Image Database Timed Test and the Enrollment Timed Test. Each of the timed tests was performed for verification and identification—once with overhead fluorescent lighting and again with the addition of back lighting.

4 How to Use This Report

The FRVT 2000 evaluations were not designed, and this report was not written, to be a buyer's guide for facial recognition. Consequently, no one should blindly open this report to a particular graph or chart to find out which system is best. Instead, the reader should study each graph and chart, the types of images used for each graph and chart, and the test method that was used to generate the graphs and charts to determine how each of them relate to the problem the reader is trying to solve. It is possible that some of the experiments performed in the Recognition Performance and Product Usability portions of this evaluation have no relation to the problem a particular reader is trying to solve and should be ignored. Once the reader has determined which image types and tests are applicable to the problem, it will be possible to study the scientific data provided and determine which system to use in a scenario and operational evaluations. The goal of this report is to provide an assessment of where the technology was in the May–June 2000 time frame. When considering face recognition technology to solve a specific problem, this report's results should be used as one of many sources to design an evaluation for your specific problem.

To understand some of the basic terms and concepts used in evaluating biometric systems, see the glossary located in Appendix N.

Table of Contents

List of Figures

List of Tables

List of Appendices

1 Introduction

1.1 Evaluation Motivation

The biggest change in the facial recognition community since the completion of the FacE REcognition Technology (FERET) program has been the introduction of facial recognition products to the commercial market. Open market competitiveness has driven numerous technological advances in automated face recognition since the FERET program and significantly lowered system costs. Today there are dozens of facial recognition systems available that have the potential to meet performance requirements for numerous applications. But which of these systems best meet the performance requirements for given applications? This is one of the questions potential users most frequently ask the sponsors and the developers of the FERET program.

Although literature research has found several examples of recent system tests, none has been both open to the public and of a large enough scale to be completely trusted. This revelation, combined with inquiries from other government agencies on the current state of facial recognition technology, prompted the DoD Counterdrug Technology Development Program Office, the Defense Advanced Research Projects Agency (DARPA), and the National Institute of Justice (NIJ) to sponsor the Facial Recognition Vendor Test (FRVT) 2000.

The sponsors decided to perform this evaluation for two main reasons. The first was to assess the capabilities of facial recognition systems that are currently available on the open market. The sponsoring agencies, as well as other government agencies, will use this information as a major factor when determining future procurement and/or development efforts. The other purpose for performing this evaluation was to show the big picture of the evaluation process and not just the results. This has numerous benefits. First, it allows others to understand the resources that would be required to run their own evaluation. Second, it sets a precedent of openness for all future evaluations. Third, it allows the community to discuss how the evaluation was performed and what modifications to the evaluation protocol could be made so that future evaluations are improved.

1.2 Qualifications for Participation

Participation in the FRVT 2000 evaluations was open to anyone selling a commercially available facial recognition system in the United States. Vendors were required to fill out forms requesting participation in the evaluation and for access to the databases used. Copies of these forms are available in Appendix A and Appendix B. Finally, the vendors were required to submit a document (maximum of four pages) that provided the following:

- An overview of the submitted system
- A component list for the submitted system
- A detailed cost breakdown of the submitted system

These documents are available in Appendix J.

Vendors were allowed to pick the components of the system, bearing in mind that results from these tests and the street price of each system at the time of testing would be made available to the public. Each vendor was allowed to submit up to two systems for testing if they could demonstrate a clear difference between the two. The final decision to allow more than one system was made by the sponsors.

2 Getting Started

2.1 Evaluation Announcement

The Facial Recognition Vendor Test 2000 was announced on February 11, 2000, by the methods described below.

- An e-mail was sent to the Biometrics Consortium (http://www.biometrics.org) listserv and directly to 24 companies that were selling facial recognition products. A copy of this e-mail announcement is provided in Appendix D.
- A description of the Facial Recognition Vendor Test 2000 was placed in the Search Biometrics area of the Counterdrug Technology Information Network (http://www.ctin.com). A copy of this posting is provided in Appendix E.

Further announcements of the evaluation were made using other means after the initial February 11 announcement date. These included:

- A success story on the FERET program was placed on the DoD Counterdrug Technology Development Program Office web site (http://www.dodcounterdrug.com). A copy of this story is provided in Appendix F.
- Links to the FRVT 2000 web site from the DARPA HumanID program web site (http://dtsn.darpa.mil/iso/programtemp.asp?mode=349)
- Included FRVT 2000 in briefings that provided an overview of the HumanID program.

2.2 Web Site

A web site for the Facial Recognition Vendor Test 2000 was created as the primary method for sharing information among vendors, sponsors and the public about the evaluation. A copy of the web site is available in Appendix C. The web site was divided into two areas—public and restricted. The public area contained the following pages.

- *Frequently Asked Questions (FAQ)*. Established to submit questions and read the responses from the evaluation sponsors.
- *Forms*. Online forms to request participation in the evaluation and for access to portions of the FERET and HumanID databases.
- *Home Page*. Menu for subsequent pages.
- *How to Participate*. Discussed how a vendor would request to participate in the evaluation.
- *Overview*. Provided the main description of the evaluation including an introduction, discussions on participant qualifications, release of the results and test make-up. This page also provided reports from the latest FERET evaluation.
- *Participating Vendors*. Provided a list of the vendors that are participating in the evaluation, a hyperlink to their web sites and point-of-contact information.
- *Points of Contact (POCs)*. Listed for test-specific questions, media inquiries and for all other questions.
- *Sponsors*. Described the various agencies that either sponsored or provided assistance for the FRVT 2000. POCs for each agency and hyperlinks to the agency's web site were provided.

- *Upcoming Dates*. Provided a list of important dates and their significance in the evaluation.

The restricted area of the FRVT 2000 web site was encrypted using 128-bit SSL encryption. Access was controlled using an ID and password provided to participating vendors and sponsors. The restricted area contained the following pages.

- *Application Programmer's Interface (API)*. Provided the application API document that shows how the vendors' similarity files would need to be written so that their results could be computed using the sponsors' scoring software. The API document was made available in both HTML and PDF formats.
- *FAQ*. This page was established to submit questions and to read the responses from the evaluation sponsors. The restricted area FAQ was more specific in nature than the public area FAQ which focused on the overview of the evaluation. See Appendix C.
- *Images*. Provided the Facial Recognition Vendor Test 2000 Demonstration Data Set, which consisted of 17 facial images in one compressed (zip) file. See Appendix I.
- *Test Plan*. Provided the detailed test plan for the evaluations. A second and final version of the test plan was also provided that answered several vendor questions about the first test plan. See Appendix H.

2.3 Conversations with Vendors

An online form was provided on the FAQ pages—public and restricted—for vendors to ask questions of the evaluation sponsors. When a form was submitted, an e-mail was automatically sent to the sponsors. The e-mail contained the submitted question and the vendor point-of-contact (POC) information for the question. A sponsor would then prepare a response, e-mail it to the vendor and post it on the FAQ web page. Some vendors preferred to use e-mail rather than the online form. When this occurred, answers were provided using the same method described above.

The practice of calling a sponsor instead of using the online form or e-mail was discouraged. Only questions of limited scope were answered via telephone, and the questions and answers were written out immediately and added to the FAQ pages for all vendors to see.

2.4 Forms

Vendors who chose to participate in the Facial Recognition Vendor Test 2000 were required to fill out two online forms from the public area of the FRVT 2000 web site—the Application for Participating in Facial Recognition Vendor Test 2000 and the Application for Access to a Portion of the Development HumanID Data Set and FERET Database. After the vendor completed all the portions of the forms and submitted them (by clicking on the submit button), three separate actions occurred. First, an e-mail, which included the field entries, was automatically sent to the evaluation sponsors. Second, this information was added automatically to a database. Third, a printer-friendly version of the form was provided to the vendors so they could print it for signature.

When a vendor submitted their online form, their information was added to the Participating Vendors page as a tentative participant. When the sponsors received the original signed copies of the form, the vendor's participation was changed to a confirmed participant. An e-mail acknkowledging receipt of the signed forms was sent to the vendor, and the vendor was given access information to the restricted area of the FRVT 2000 web site.

2.5 Time Line

The Facial Recognition Vendor Test 2000 was announced on February 11, 2000. The final day for vendors to sign up was March 17, 2000. On this date, eight vendors had requested and been approved to participate in the evaluation. Two others had also inquired about participating but did not sign up.

An Image Development set and an API document for a portion of the evaluation were released on March 8. On March 27, vendors submitted sample similarity files based on the Image Development set and the API document so the sponsors could test their compliance. A few vendors had errors in their similarity files and had to resubmit modified similarity files. All vendors eventually submitted correct similarity files and were notified of this on April 3.

The test schedule and detailed test plan were released on March 27. On March 31 a revised version was released that clarified some areas in response to participating vendors' questions and lessons learned from practice sessions with the test subjects.

On March 20, one of the eight participating vendors withdrew from the evaluation stating, "[We] have concluded that the Vendor Test 2000 is too unconstrained for our currently released product. Although we are very close to releasing our auto head detection and head rotation product for unconstrained environments, we feel it is a bit premature since it has not undergone rigorous field testing yet." On March 21, two more participating vendors withdrew from the evaluation. One vendor cited a difference of opinion on how the systems were to be evaluated in FRVT 2000, and the other gave no reason for their withdrawal. On March 22, a fourth participating vendor withdrew from the evaluation, citing a need to allocate their resources to a government contract that had several deliverables due at the time the evaluations were to take place. Subsequently, this vendor requested reinstatement and was accepted (with a new point of contact) on March 28. This left five participating vendors.

Each vendor had a full week to perform the test. Some vendors provided preferred dates for their test, and each was given their first choice. Foreign vendors were deliberately placed last on the test schedule because they needed extra time to work with their embassies to obtain access to NAVSEA Crane. Each vendor was allowed to choose which day of their test week to schedule each of the subtests discussed in Section 4.1. The final schedule is shown below.

- May 1–5—Visionics Corp.
- May 8–12—Lau Technologies
- May 15–19—Miros Inc. (eTrue)
- May 22–26—C-VIS Computer Vision und Automation GmbH
- June 5–9—Banque-Tec International Pty. Ltd.

3 Writing the Evaluation Methodology

3.1 Background

The sponsors of the Facial Recognition Vendor Test 2000 talked with numerous government agencies and several members of the biometrics community, including facial recognition vendors, to determine if this evaluation should be and how it would be performed. The overwhelming response was to proceed with the evaluation. Government agencies and the biometrics community wanted to

know if the facial recognition vendors could live up to their claims, which systems performed best in certain situations and what further development efforts would be needed to advance the state of the art for other applications. Unofficially, the vendors wanted to have an evaluation to prove that *they* had the best available product. Everyone cited the FERET program because it is the *de facto* standard for evaluating facial recognition systems, but they also stressed the need to have a live evaluation.

FRVT 2000 sponsors took this information and began analyzing different methods to evaluate facial recognition systems. Three items had a profound effect on the development of the FRVT 2000 evaluation methodology:

- "An Introduction to Evaluating Biometric Systems," P. J. Phillips, A. Martin, C. L. Wilson, M. Przybocki, *IEEE Computer,* February 2000, p. 56–63.
- The FERET program.
- A previous scenario evaluation of a COTS facial recognition system.

3.2 An Introduction to Evaluating Biometric Systems

The FRVT 2000 sponsors received an early draft of the article written by P. Jonathon Phillips, et al, and also reviewed a later draft before publication. Numerous ideas were taken from this paper and used in the FRVT 2000 evaluations.

The first idea taken was that the evaluations should be administered by independent groups and tested on biometric signatures not previously seen by a system. The sponsors of the FRVT 2000 felt that these two items were necessary to ensure the integrity of the evaluation and its results. Another idea was that the details of the evaluation procedure must be published along with the evaluation protocol, testing procedures, performance results and representative examples of the data set. This would ensure that others could repeat the evaluations. An evaluation must also not be too difficult or too easy. In either case, results from varying vendors would be grouped together and a distinction between them would not be possible. This is depicted in figure 1[1].

The final idea taken from this paper was the concept of a three-step evaluation plan: a technology evaluation, a scenario evaluation and an operational evaluation. The goal of the technology evaluation was to compare competing algorithms from a single technology—in this case facial recognition. Algorithm testing is performed on a standardized database collected by a universal sensor—the same images are used as input for each system. The test should also be performed by an organization that will not benefit should one algorithm outperform the others. Using a test set ensures that all participants see the same data. Someone who is interested in facial recognition can look at the results from the image sets that most closely resemble their situation and determine, to a reasonable extent, what results they should expect. At this point potential users can develop a scenario evaluation based on their real-world application of interest and invite selected systems to be tested against this scenario. Each tested system would have its own acquisition sensor and would receive slightly different data. The application that performs best in the scenario evaluation can then be taken to the actual site for an extended operational evaluation before purchasing a complete system. This three-step evaluation plan has also been adopted by Great Britain's Best Practices in Testing and Reporting Performance of Biometric Devices. This report can be found at http://www.afb.org.uk/bwg/bestprac10.pdf.

[1] P. J. Phillips, H. Moon, P. J. Rauss, S. Risvi, "The FERET Evauluation Methodology for Face Recognition Algorithms," *IEEE Trans Pattern Analysis and Machine Intelligence*, Vol. 22, No. 11, p. 1090–1104, 2000.

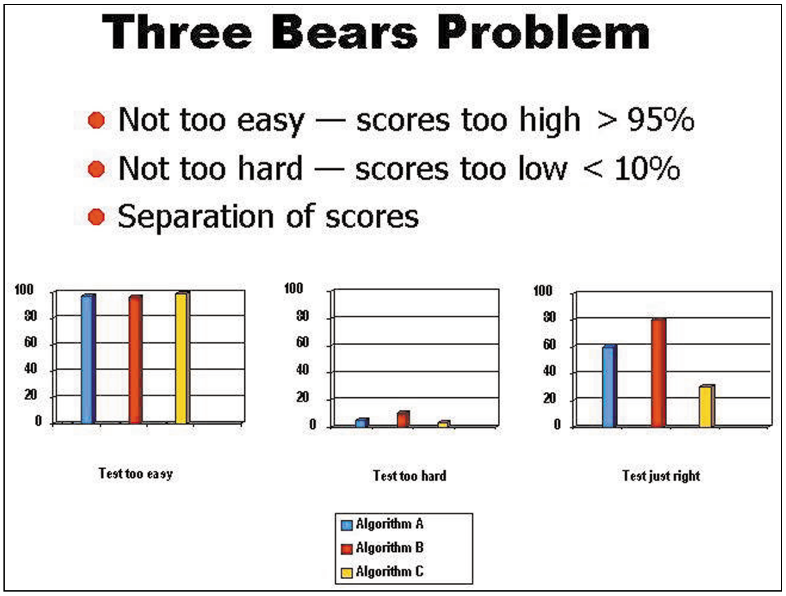

Figure 1: *Three Bears Problem*

3.3 The FERET Program

The DoD Counterdrug Technology Development Program Office began the FacE REcognition Technology (FERET) program in 1993. The program consists of three important parts:

- Sponsoring research.
- Collecting the FERET database.
- The FERET evaluations.

FERET-sponsored research was instrumental in moving facial recognition algorithms from concept to reality. Many commercial systems still use concepts that were involved in the FERET program as seen in figure 2.

The FERET database was designed to advance the state of the art in facial recognition, with the images collected directly supporting algorithm development and the FERET evaluations. The database is divided into a development set, which was provided to researchers, and a set of images that was sequestered. The sequestering was necessary so that additional FERET evaluations and future evaluations such as the FRVT 2000 could be administered using images that researchers have not previously used with their systems. If previously used images are used in an evaluation, it is possible that researchers may tune their algorithms to handle that specific set of images. The FERET database contains 14,126 facial images of 1,199 individuals. Before the FRVT 2000, only one-third of the FERET database had ever been used by anyone outside the government. The DoD Counterdrug Technology Development Program Office still receives requests for access to the FERET database, which is maintained at the National Institute of Standards and Technology (NIST). The FERET development set has been distributed to more than 100 groups outside the original FERET program.

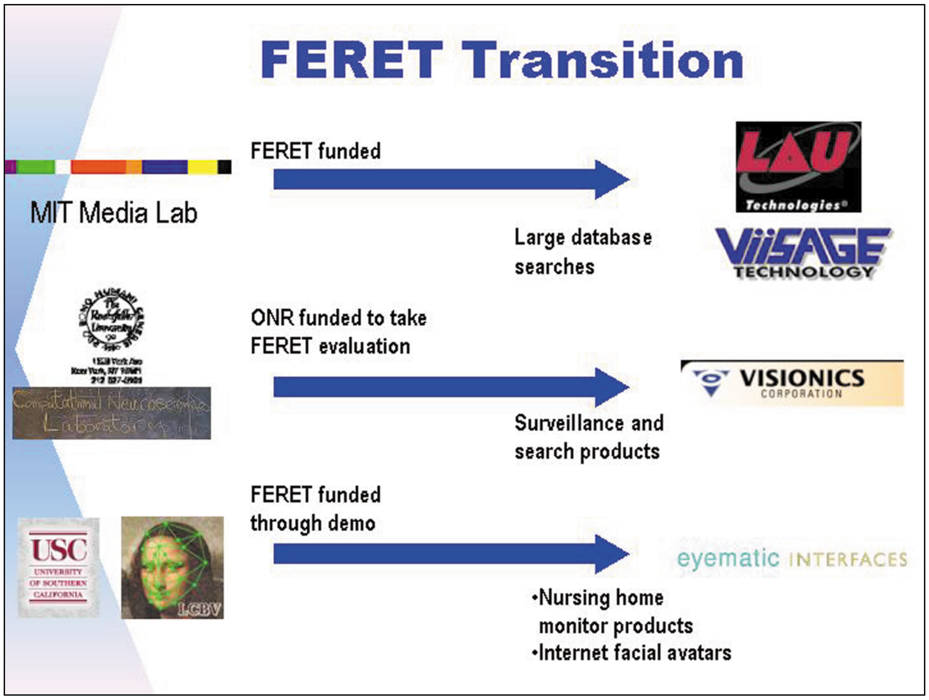

Figure 2: *FERET Transition*

The final and most recognized part of the FERET program was the FERET evaluation[2] that compared the abilities of facial recognition algorithms using the FERET database[3]. Three sets of evaluations were performed in August 1994, March 1995 and September 1996.

A portion of the FRVT 2000 has been based very heavily on the FERET evaluation. Numerous images from the unreleased portion of the FERET database, the scoring software and baseline facial recognition algorithms for comparison purposes were used in FRVT 2000. The FERET program also provided insight into what the sponsors should expect from participants and outside entities before, during and after the evaluations.

3.4 A Previous Scenario Evaluation for a COTS Facial Recognition System

In 1998, the DoD Counterdrug Technology Development Program Office was asked to study the feasibility of using facial recognition at an access control point in a federal building. The technical agents assigned from NAVSEA Crane Division studied the layout and arranged a scenario evaluation for a facial recognition vendor at their facilities. The selected vendor brought a demonstration system to NAVSEA Crane, set it up and taught the technical agents how to use the system.

A subject was enrolled into the system according to the procedures outlined by the vendor. During the evaluation, the technical agent entered the subject's ID number into the system, which was configured for access control (verification) mode. A stopwatch was used to measure the recognition time starting with the moment the ID number was entered and ending when the subject was

[2] P. J. Phillips, H. Moon, P. J. Rauss, S. Risvi, "The FERET Evauluation Methodology for Face Recognition Algorithms," *IEEE Trans Pattern Analysis and Machine Intelligence*, Vol. 22, No. 11, p. 1090–1104, 2000.

[3] P. J. Phillips, H. Wechsler, J. Huang, P. Rauss, "The FERET Database and Evaluation Procedure for Face Recognition Algorithms," *Image and Vision Computing Journal*, Vol. 16, No. 5, p. 295–306, 1998.

correctly identified by the system. The resulting time, measured in seconds, was recorded in a table. This timed test was repeated at several distances with the subject being cooperative and indifferent. System parameters were also varied incrementally from one extreme to the other. The methodology of the evaluation was never explained to the vendor.

When the system was returned to the vendor, they looked at the system settings for the final iteration of the timed test and immediately complained that NAVSEA Crane had not tested the system at an optimal point. They offered to return to NAVSEA Crane with another system so they could retest using the vendor's own test data and test plan and then write a report that the sponsors could use instead of the sponsor-written evaluation report. The invitation was not accepted because the proposed effort had been canceled for other reasons.

The DoD Counterdrug Technology Development Program Office learned several lessons from this simple evaluation. The first was how to develop a scenario evaluation and improve on it for future evaluations such as the FRVT 2000. The second lesson was the importance of being completely candid about the evaluation plan so the vendor is less inclined to dispute its validity after the evaluation. The final and most important lesson was to continue to let a non-biased sponsor run the evaluations, but allow a vendor representative to run their own machines and set the system parameters under the sponsor's supervision. Because the sponsor, rather than the vendor representative, ran the system during the evaluation, this gave the vendor an opportunity to blame poor results on operator error rather than the system.

All three lessons were used to develop the evaluation methodology for the FRVT 2000.

4 FRVT 2000 Description

4.1 Overview

The Facial Recognition Vendor Test 2000 was divided into two evaluation steps: the Recognition Performance Test and the Product Usability Test. The FRVT 2000 Recognition Performance Test is a technology evaluation of commercially available facial recognition systems. The FRVT 2000 Product Usability Test is an example of a scenario evaluation, albeit a limited one.

After completing the evaluation, all test images, templates, and similarity files were deleted from the vendor machine and all hard disk free space was wiped. Vendors then signed forms stating that the data recorded for the Product Usability Test were accurate, and they would not share the data with anyone outside their organization until after the results were publicly released by the sponsors. Vendors were given copies of these signed forms as well as the completed data recording tables.

4.2 Test Procedures

The test was run according to the test plan provided to vendors before testing began. A copy of the test plan is included in Appendix H.

As testing started with the first vendor, a few minor adjustments were made to the procedures and applied consistently for each vendor test. The original plan was to use subject 3 for the variability test. The range of subject heights, however, made it difficult to adjust the camera so that all subjects would be in the field of view at very close range. The bottom of the face was sometimes out of range for the shortest subject and the top of the face for the tallest subject. It was decided to use subject 1, who was in between the height extremes, as the subject for the variability test because he was always in view at close range. Originally, it was decided that acquire times would be recorded to the nearest

1/10 second for the Product Usability Test. The stopwatch used for the test, however, displayed time in 1/100 of a second increments. The decision was made to record the times to the nearest 1/100 second rather than round or truncate the displayed time.

5 Evaluation Preparations

5.1 Image Collection and Archival

Image collection and archival are two of the most important aspects of any evaluation. Unfortunately, they do not normally receive enough attention during the planning stages of an evaluation and are rarely mentioned in evaluation reports. Without a very controlled (or purposely uncontrolled) image collection protocol that is released with the evaluation results, no one would understand what the results mean. For example, vendor A can point to results from one database subset and vendor B can point to different results. It is impossible to make an accurate assessment of capabilities from this comparison, but it is routinely done. Another example is to provide results from an independent analysis where each vendor was compared using the same database subset. This is a better practice, but as the results section of this report will demonstrate, wide variations can occur based on the types of images used. Unless a description of the image collection process is included with the results, the validity of any conclusions from those tests is questionable.

The Facial Recognition Vendor Test 2000 used images from the FERET database and the HumanID database. The FERET database has been discussed in previous reports. The portion of the HumanID database used in FRVT 2000 was collected by the National Institute of Standards and Technology. A description of the collection setup, processing and post-processing performed by NIST is provided in Appendix G.

5.2 Similarity File Check

The sponsors of FRVT 2000 wanted to make sure that the output produced by vendor software during the Recognition Performance Test could be read successfully and processed by the sponsor-developed scoring software. The goal was to resolve any potential problems before testing began. Participating vendors were required to compare each of the 18 images in the Image Development set with each of the other images in the set and create similarity files according to the format described in the API document. These similarity files were e-mailed to the sponsors for compliance verification. The software tried to read each of the ASCII files containing similarity scores and returned error messages if any compliance problems were found. A few vendors had errors in their similarity files and were asked to resubmit modified similarity files. All participating vendors eventually submitted correct similarity files and were notified of this.

5.3 Room Preparation

Several weeks before the tests began, the testing room was prepared. The arrangement of the different test stations is described in Appendix H. Figures 3 and 4 show a detailed layout of the room and the locations of the overhead fluorescent lights.

5.4 Backlighting

Backlighting was used for some trials in the timed tests. This was to simulate the presence of an outside window behind the subject in a controlled and repeatable manner. To accomplish this, a custom lighting device was built. It consists of a track lighting system with fixtures arranged in a 4 x 4 grid. The lights used for this device were manufactured by Solux and chosen because they have a spec-

tral power distribution that closely mimics that of daylight. The particular model used for this application has a beam spread of 36 degrees and a correlated color temperature of 4,700 degrees Kelvin. Power requirements for each bulb are 50 watts at 12 volts. The 4 x 4 light grid was mounted inside a box facing toward the camera. The inside of the box was covered with flat white paint. The front side of the box, which faced the camera, was 4 ft. x 4 ft. The material used on the front side is a Bogen Lightform P42 translucent diffuser panel. The lights were arranged so the beams overlapped on the surface of the front panel for even illumination.

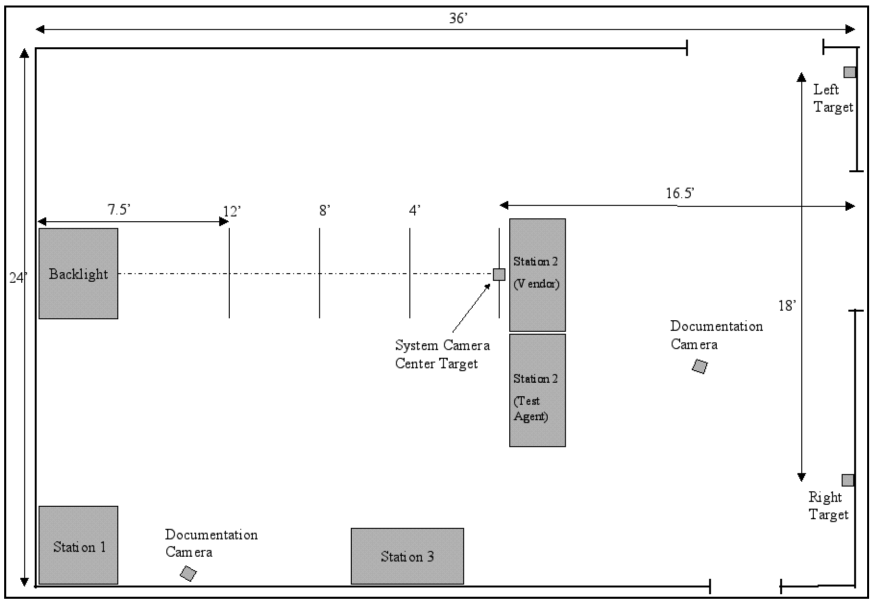

Figure 3: *Testing room layout*

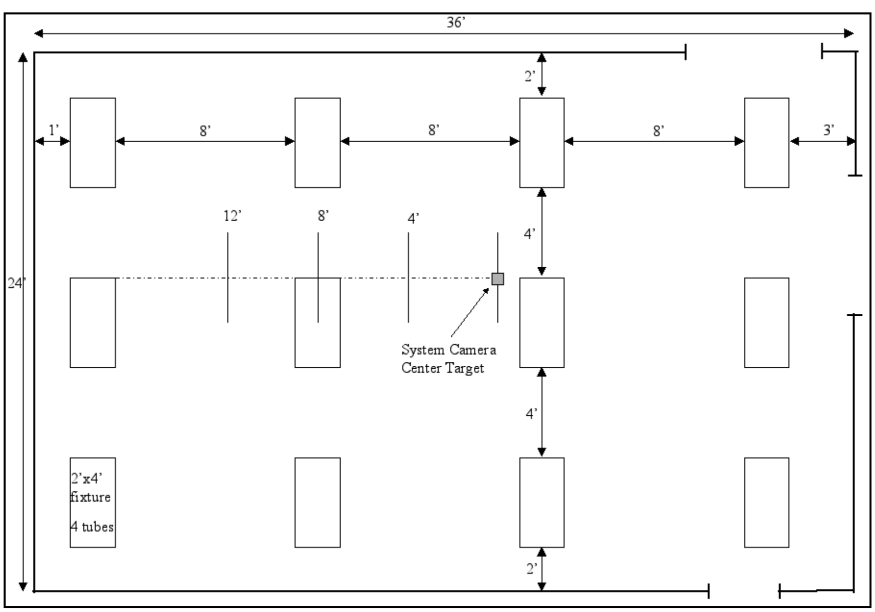

Figure 4: *Fluorescent light layout for testing room*

5.5 Subject Training

In the weeks leading to the first test date, the test agent met several times with the three test subjects in the room where the testing would take place. The purpose of these meetings was to explain the Product Usability Test procedures described in the test plan, let the subjects practice their roles to achieve consistent behavior before the tests began and uncover any problems with the test plan procedures. The subjects practiced walking in front of a camera about 15 times each at the first meeting. During this session, a few procedural improvements were suggested and implemented by the subjects.

- Use a metronome set to 60 beats per minute to synchronize walking cadence and head movement, giving more consistent results with each trial.
- Draw more attention to the stop marker placed one foot in front of the camera so the subjects could more easily detect this location while walking and turning their heads during the indifferent trials.
- Begin identification trials with bodies one-quarter turned from the camera path to help ease the awkwardness of the 180-degree turn specified in the original test plan.

To accomplish these improvements, a metronome was purchased. Two tripods were placed at the stop marker with yellow caution tape stretched between them at a height of 3 feet for added visibility using peripheral vision. The test plan was updated to specify facing 90 degrees from the camera path at the beginning of identification trials.

After the improvements were made and the test procedures were updated, two more practice sessions were held. Each session lasted approximately one hour, and each subject participated in about 20 to 25 trials. Both sessions were held the week before the first vendor test to keep the procedures fresh in the subjects' minds.

5.6 Scoring Algorithm Modification

The similarity file scoring algorithm, used for the Recognition Performance portion of the FRVT 2000 evaluations, was originally developed for the FERET program. After the FERET program concluded, NIJ and DARPA cofunded an update to the algorithm so it can use the C/C++ programming language and a revised ground-truth format. The scoring algorithm was updated again for the FRVT 2000 evaluations so it could function with a less than complete set of similarity files. The new scoring algorithm was validated using three different methods.

The first validation method used the baseline PCA algorithm developed for the FERET program to develop similarity files using the same set of images used in the September 1996 FERET evaluations. The images were then scored using the new scoring algorithm and the resulting CMC curves (see Section 7.1.2) were compared to the original results.

The second validation method the sponsors used was to write an algorithm that synthesizes a set of similarity files from a given CMC curve. The new scoring algorithm then scored the similarity files and the results were compared to the original curve for validation.

The third validation method was to provide the participating vendors with a set of similarity files derived from a baseline algorithm using FERET images, the scoring software and the results from the scoring software. Participating vendors were then asked to study the validity of the scoring code and provide feedback to the evaluation sponsors if they found any software implementation errors. The vendors did not report any errors.

6 Modifications

During the course of the evaluation, the original plan had to be modified to accommodate events that occurred. The minor modifications have been discussed in previous chapters. The following sections outline the other modifications and the reasoning behind them.

6.1 Access Control System Interface Test

Only one vendor opted to take the access control system interface test, which was part of the Product Usability Test. During the test, it was noted that there was not enough information available about the access control system to make a proper signal connection with the vendor system. Some proprietary details were needed that could not be obtained within the time allowed for the test. To connect the systems, the facial recognition vendor needed to obtain details on the WIEGAND interface from the access control vendor. Since the WIEGAND protocol has many parameters that vary between systems, the facial recognition system could not be connected to the access control system without custom configuration. As a result, the Access Control System Interface Test was abandoned and no further results will be published in this report. Our conclusion is that anyone who wants to connect a facial recognition system to an access control system at this time should expect the process to include some custom development work.

6.2 FERET Images

Three of the major objectives of the Facial Recognition Vendor Test 2000 were to provide a comparison of commercially available systems, provide an overall assessment of the state of the art in facial recognition technology and measure progress made since the conclusion of the FERET program.

The comparison of commercially available systems needed to be designed and administered so that all vendors were on a level playing field and inadvertent advantages were not given to any participants. One of the methods used to ensure this in FRVT 2000 was to administer the test using sequestered images from the FERET program that had not been included in any previous evaluations. Any image set that was established for testing, however, has a certain life cycle associated with it. Once it has been used extensively and results using the data set have been published, developers start to learn the properties of the database and can begin to game or tune their algorithms for the test. This is certainly true of the FERET database; portions of it have been used in evaluations since August 1994. The FERET database has also been used in numerous other studies. To ensure a fair and just evaluation of the commercial systems in FRVT 2000, individual results for each vendor will be given using only those images that had been collected since the last FERET evaluations.

Another objective of the FRVT 2000 was to provide the community a way to assess the progress made in facial recognition since the FERET program concluded. There are two ways to measure progress. The best is to have the algorithms used in previous evaluations subjected to the new evaluation. Unfortunately, this was not an option for the FRVT 2000. The next best solution is to have the previous evaluation included in the current evaluation. This appears to be at odds with the goal of having an unbiased evaluation because those who participated in previous evaluations would have an advantage over those who did not. Because the goal is to measure progress and not necessarily individual system results, we can work around the potential conflict by reporting the top aggregate score from the experiments that used the FERET database.

The third goal—an overall assessment of the state of the art in facial recognition technology—

can be inferred by looking at the combined results from the commercial system evaluation and the results using the FERET data.

6.3 Reporting the Results

For the Recognition Performance portion of this evaluation, the vendors were asked to compare 13,872 images to one another, which amounts to more than 192 million comparisons. The vendors were given 72 continuous hours to make these comparisons and then told to stop making their comparisons. C-VIS, Lau Technologies and Visionics Corp. successfully completed the comparison task. Banque-Tec completed approximately 9,000 images, and Miros Inc. (eTrue) completed approximately 4,000 images in the time allowed.

The complete set of 13,872 images and the corresponding matrix of 13,872 x 13,872 similarity scores can be divided into several subsets that can be used as probe and gallery images for various experiments. Probe images are presented to a facial recognition system for comparison with previously enrolled images. The gallery is the set of known images enrolled in the system.

Banque-Tec and Miros Inc. (eTrue) completed only a small number of the FRVT 2000 experiments and submitted only partial responses to several more. This forced the evaluation sponsors to decide how to accurately provide results from the FRVT 2000 experiments. The following options were considered.

Option 1 was to only release the results from the experiment that all five vendors completed (M2). This was rejected because this one experiment does not adequately describe the current capabilities of the commercial systems.

Option 2 was to release results from all of the FRVT 2000 experiments and only show the results from the vendors that completed each experiment. This would show the results for C-VIS, Lau Technologies and Visionics Corp. for all experiments and add the results from Banque-Tec and Miros Inc. (eTrue) for the M2 experiment. The sponsors chose not to do this because of the possibility that these two vendors may have received an added advantage in this category because they took more time to make the comparisons. Although the data collected does not support this hypothesis, the sponsors felt it would be better to not allow this argument to enter the community's discussion of the FRVT 2000 evaluations.

Option 3 was to change the protocol of the experiments so, for example, the D3 category only used the probes that all five vendors completed rather than the entire set. This option was rejected for the same reasons stated in Option 2.

Option 4 was to show the results from C-VIS, Lau Technologies and Visionics Corp. based on the full probe sets for each experiment and the results from Banque-Tec and Miros Inc. (eTrue) based on the subset that they completed. This option was rejected for the same reason stated in Option 2.

Option 5 was to fill the holes in the similarity matrices of Banque-Tec and Miros Inc. (eTrue) with a random similarity score or the worst similarity score that they had provided to that point. This option was rejected because the results generated would be horrendous and significantly skew the results that had been provided.

Option 6 was to show the results from C-VIS, Lau Technologies and Visionics Corp. and ignore the results from Banque-Tec and Miros Inc. (eTrue) for the FRVT 2000 experiments. This option was selected because it was the only one that was fair and just to those that had finished the required number of images and those that had not.

7 FRVT 2000 Results

7.1 Recognition Performance Test

7.1.1 Overview

Each vendor was given a set of 13,872 images to process. They were instructed to compare each image with itself and with all other images, and return a matching score for each comparison. The matching scores were stored in similarity files that were returned to the test agent along with the original images. Each vendor was given 72 continuous hours to process the images. Some vendors were able to process the entire set of images, while others were only able to process a subset of the images in the allotted time. At the conclusion of the test, each vendor's hard disk was wiped to eliminate the images, similarity files and any intermediate files.

After all testing activities were complete, the similarity files were processed using the scoring software. The images were divided into different probe and gallery sets to test performance for various parameters such as lighting, pose, expression and temporal variation. The results for each of these probe and gallery sets are reported here in bar charts that highlight key results. The full receiver operator characteristic (ROC) and cumulative match characteristic (CMC) for each experiment are shown in Appendix M.

7.1.2 Interpreting the Results – What Do the Charts Mean?

Biometric developers and vendors will, in many cases, quote a false acceptance rate (sometimes referred to as the false alarm rate) and a false reject rate. A false acceptance (or alarm) rate (FAR) is the percentage of imposters (an imposter may be trying to defeat the system or may inadvertently be an imposter) wrongly matched. A false rejection rate (FRR) is the percentage of valid users wrongly rejected. In most cases, the numbers quoted are quite extraordinary. They are, however, only telling part of the story.

The false acceptance rate and false rejection rate are not mutually exclusive. Instead, there is a give-take relationship. The system parameters can be changed to receive a lower false acceptance rate, but this also raises the false rejection rate and vice versa. A plot of numerous false acceptance rate-false rejection rate combinations is called a receiver operator characteristic curve. A generic ROC curve is shown in figure 5. The probability of verification on the y-axis ranges from zero to one and is equal to one minus the false reject rate. The false acceptance (or alarm) rate and the false reject rate quoted by the vendors could fall anywhere on this curve and are not necessarily each other's accompanying rate. Some spec sheets also list an equal error rate (EER). This is simply the location on the curve where the false acceptance rate and the false reject rate are equal. A low EER can indicate better performance if one wants to keep the FAR equal to the FRR, but many applications naturally prefer a FAR/FRR combination that is closer to the end points of the ROC curve. Rather than using EER alone to determine the best system for a particular purpose, one should use the entire ROC curve to determine the system that performs best at the desired operating location. The ROC curve shown in figure 5 uses a linear axis to easily show how the equal error rate corresponds to the false acceptance and false reject rate. The ROC curves in Appendix M that show actual FRVT 2000 results use a semi-log axis so that low-false-alarm rate results can be viewed. The equal error rates are listed as text on the graphs.

Although an ROC curve shows more of the story than a quote of particular rates, it will be difficult to have a good understanding of the system capabilities unless one knows what data was used to make these curves. An ROC curve for a fingerprint system that obtained data from coal miners would

be significantly different than one that obtained data from office workers. Facial recognition systems differ in the same way. Lighting, camera types, background information, aging and other factors would each impact a facial recognition system's ROC curve. For the Facial Recognition Vendor Test 2000, participating vendors compared 13,872 images to one another. These images can be subdivided into different experiments to make an ROC curve that shows the results of comparing one type of image to another type of image. Section 7.1.3 describes the different experiments that will be reported.

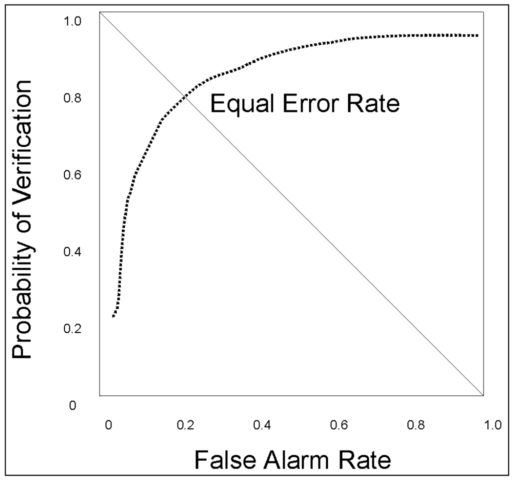

Figure 5: *Sample Receiver Operating Characteristic (ROC) with an EER of 0.2*

The above description is valid for displaying verification results. In a verification application, a user claims an identity and provides their biometric. The biometric system compares the biometric template (the digital representation of the user's distinct biometric characteristics) with the user's stored (upon previous enrollment) template and gives a match or no-match decision. Biometric systems can also act in an identification mode, where a user does not claim an identity but only provides their biometric. The biometric system then compares this biometric template with all of the stored templates in the database and produces a similarity score for each of the stored templates. The template with the best similarity score is the system's best guess at who this person is. The score for this template is known as the top match.

It is unrealistic to assume that a biometric system can determine the exact identity of an individual out of a large database. The system's chances of returning the correct result increases if it is allowed to return the best two similarity scores, and increased even more if it is allowed to return the best three similarity scores. A plot of probabilities of correct match versus the number of best similarity scores is called a cumulative match characteristics curve. A generic CMC curve is shown in figure 6.

Just as with ROC curves, these results can vary wildly based on the data that was used by the biometric system. Results for the same experiments described in Section 7.1.3 for verification results will also be shown for identification results. One other item must be provided to complete the story for CMC results: the number of biometric templates in the system database. This number is also provided in Section 7.1.3.

The ROC and CMC curves that show each vendor's results for the experiments defined in Section 7.1.3 are located in Appendix M. The sponsors found it difficult to quickly compare results between experiments and vendors using the ROC and CMC curves. Key points of these results are shown in Section 7.1.3 in the form of bar charts.

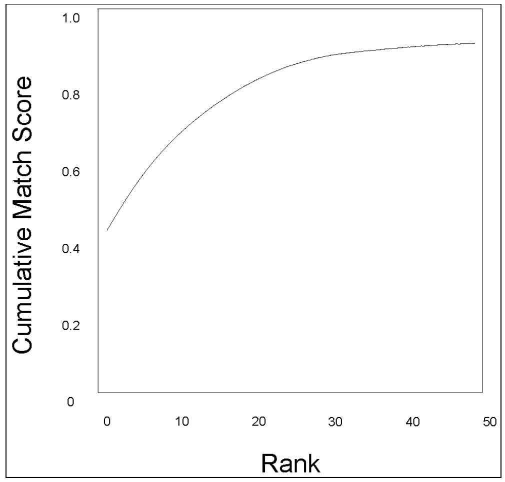

Figure 6: *Sample Cumulative Match Characteristic (CMC)*

7.1.3 Recognition Performance Test Experiment Descriptions

Numerous experiments can be performed based on the similarity files returned by the participating vendors. The following subsections, along with tables 1–9, describe the experiments performed by the sponsors for this report. The rows with a white background are designated as FRVT 2000 experiments, while the rows with a gray background are designated as FERET experiments.

To make comparisons between vendors and between experiments easier, the sponsors have highlighted key results via bar charts in figures 7-63. The complete ROC and CMC curves are located in Appendix M and should be studied to gain a complete understanding of the systems' capabilities.

Results shown in this section are from experiments that use images from the FERET database. The purpose of these experiments is to assess the improvement made in the facial recognition community since the conclusion of the FERET program. Results for individual vendors are not given for these experiments. Rather, the sponsors developed best CMC curves by choosing the top score at each rank from the results obtained from C-VIS, Lau Technologies and Visionics Corp. See Section 7.1.2 for a detailed explanation of CMC curves.

Table 1: *List of experimental studies reported, tables describing experiments, figures and page numbers for reported results, and names of experiments in each study.*

Experiment Name	Experiment Study	Table Number	Figure Numbers	Start Page
C0–C4	Compression	2	7, M–1	17
D1–D7	Distance	3	8, M–12, M–34	19
E1–E2	Expression	4	26, M–19, M–41	25
I1–I3	Illumination	5	32, M–21, M–43	28
M1–M2	Media	6	38, M–24, M–46	30
P1–P5	Pose	7	44, M–6, M–26, M–48	33
R1–R4	Resolution	8	51, M–27, M–49	37
T1–T5	Temporal	9	57, M–10, M–31, M–53	41

7.1.3.1 Compression Experiments

The compression experiments were designed to estimate the effect of lossy image compression on the performance of face-matching algorithms. Although image compression is widely used to satisfy space and bandwidth constraints, its effect in machine vision applications is often assumed to be deleterious; therefore, compression is avoided. This study mimics a situation in which the gallery images were obtained under favorable, uncompressed circumstances, but the probe sets were obtained in a less favorable environment in which compression has been applied. The amount of compression is specified by the compression ratio. The probe sets contain images that were obtained by setting an appropriate quality value on the JPEG compressor such that the output is smaller than the uncompressed input by a factor equal to the compression ratio.

The imagery used in these experiments is part of the FERET corpus; the native source format is uncompressed. The gallery used for the compression experiments is the standard 1,196-image FERET gallery. The probe set used is the 722 images from the FERET duplicate I study.

Table 2: *Figures showing results of JPEG compression experiments. Gallery and probe images were generated from the T1 (Dup 1) study. All images are from the FERET database.*

Experiment Name	Figure Numbers	Compression Ratio	Gallery Size	Probe Set Size
C0	7, M–1	1:1 (none)	1,196	722
C1	7, M–2	10:1	1,196	722
C2	7, M–3	20:1	1,196	722
C3	7, M–4	30:1	1,196	722
C4	7, M–5	40:1	1,196	722

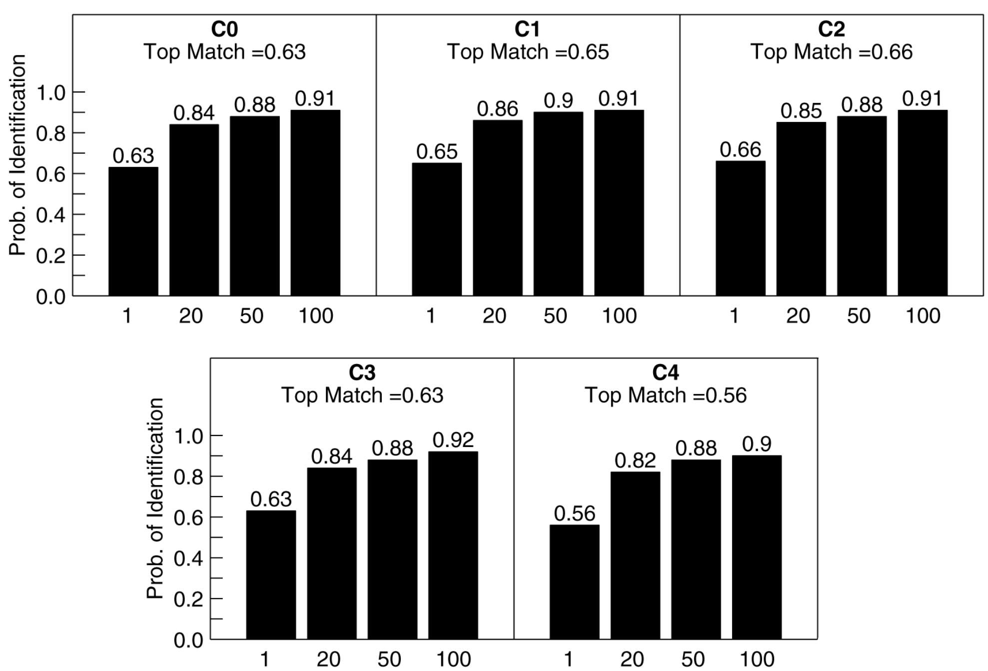

Figure 7: *FERET Results—Compression Experiments Best Identification Scores*

7.1.3.2 Distance Experiments

The distance experiments were designed to evaluate the performance of face matching algorithms on images of subjects at different distances to the fixed camera. The results of these experiments should be considered for situations where the distance from the subject to the camera for enrollment is different from that used for verification or identification.

In all experiments, the probe images were frames taken from relatively low-resolution, lightly compressed, video sequences obtained using a consumer grade tripod-mounted auto-focus camcorder. In these sequences the subjects walked down a hallway toward the camera. Overhead fluorescent lights were spaced at regular intervals in the hallway, so the illumination changed between frames in the video sequence. This may be thought of as mimicking a low-end video surveillance scenario such as that widely deployed in building lobbies and convenience stores. Two kinds of galleries were used: In experiments D1-D3 the gallery contains images of individuals with normal facial expressions that were acquired indoors using a digital camera under overhead room lights. In experiments D4-D7, however, the gallery itself contains frames extracted from the same video sequences used in the probe sets. Experiments D1-D3, therefore, represent a mugshot vs. subsequent video surveillance scenario in which high-quality imagery is used to populate a database and recognition is performed on images of individuals acquired on video. Experiments D4-D7 test only the effect of distance and avoid the variation due to the camera change.

Note that although the study examines the effect of increasing distance (quoted approximately in meters) the variable often considered relevant to face recognition algorithms is the number of pixels on the face. The distance and this resolution parameter are inversely related. The resolution studies described later also address this effect.

The D4-D5 and D6-D7 studies may be compared to provide a qualitative estimate to the effect of indoor and outdoor lighting. This aspect is covered more fully in the illumination experiments that follow.

Table 3: *Figures showing results of distance experiments. All images are from the HumanID database, and all gallery and probe images are frontal.*

		Gallery Images		Probe Images			
Experiment Name	**Figure Numbers**	**Description**	**Camera Distance**	**Description**	**Camera Distance**	**Gallery Size**	**Probe Set Size**
D1	8, 11, 14, 17, 20, 23, M–12, M–34	Indoor, digital, ambient lighting	1.5 m	Indoor, video	2 m	185	189
D2	8, 11, 14, 17, 20, 23, M–13, M–35	Indoor, digital, ambient lighting	1.5 m	Indoor, video	3 m	185	189
D3	8, 11, 14, 17, 20, 23, M–14, M–36	Indoor, digital, ambient lighting	1.5 m	Indoor, video	5 m	185	189
D4	9, 12, 15, 18, 21, 24, M–15, M–37	Indoor, video	2 m	Indoor, video	3 m	182	190
D5	9, 12, 15, 18, 21, 24, M–16, M–38	Indoor, video	2 m	Indoor, video	5 m	182	190
D6	10, 13, 16, 19, 22, 25, M–17, M–39	Outdoor, video	2 m	Outdoor, video	3 m	186	195
D7	10, 13, 16, 19, 22, 25, M–18, M–40	Outdoor, video	2 m	Outdoor, video	5 m	186	195

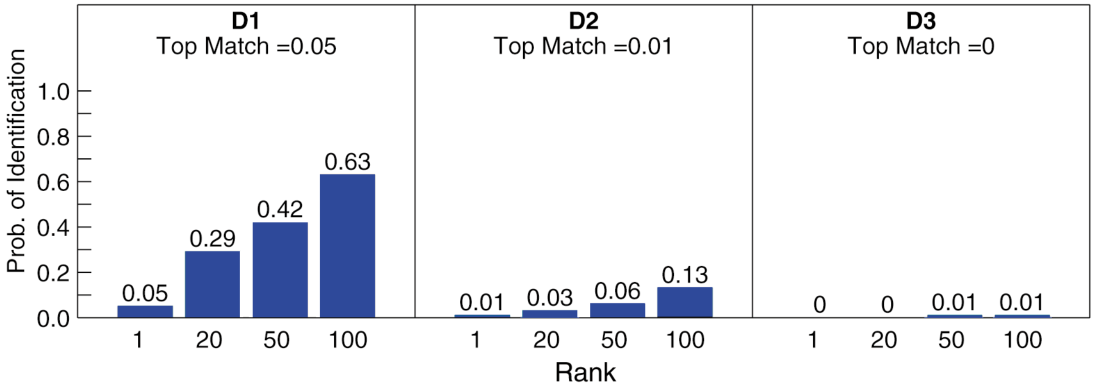

Figure 8: *FRVT 2000 Distance Experiments—C-VIS Identification Scores*

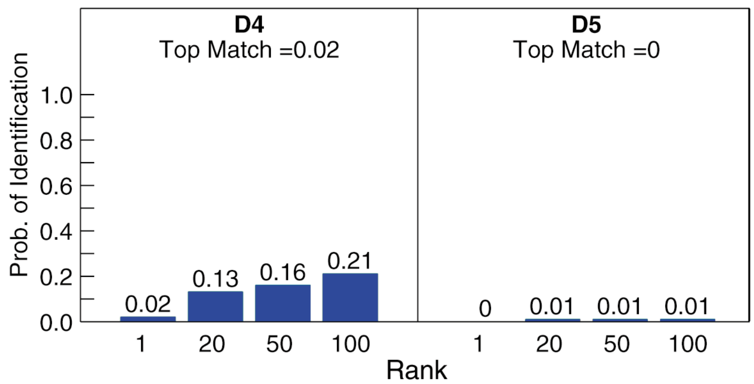

Figure 9: *FRVT 2000 Distance Experiments—C-VIS Identification Scores*

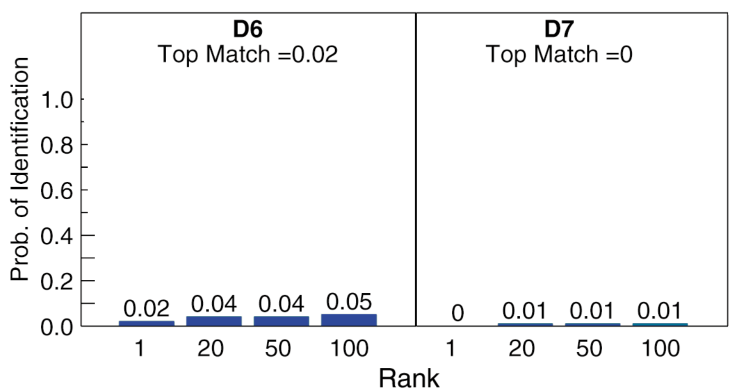

Figure 10: *FRVT 2000 Distance Experiments—C-VIS Identification Scores*

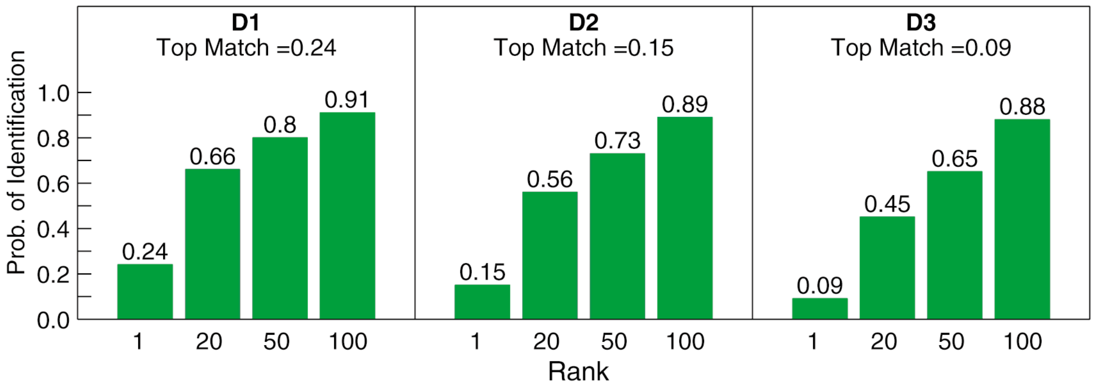

Figure 11: *FRVT 2000 Distance Experiments—Lau Technologies Identification Scores*

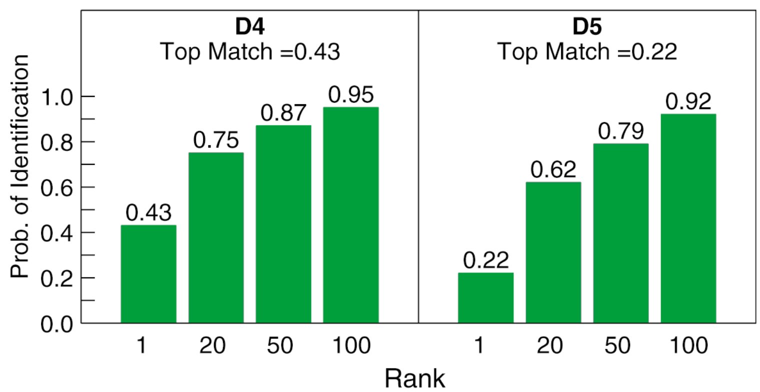

Figure 12: *FRVT 2000 Distance Experiments—Lau Technologies Identification Scores*

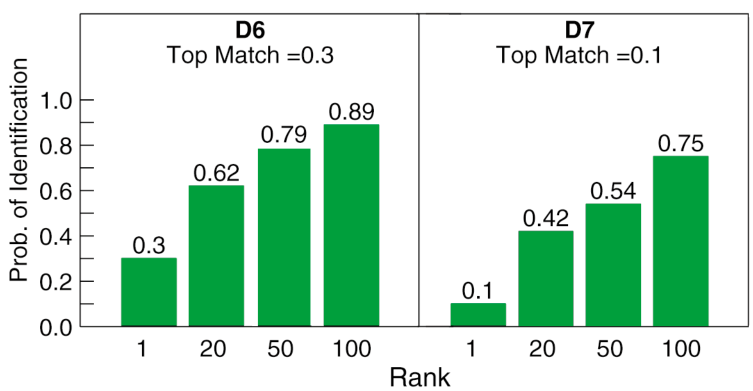

Figure 13: *FRVT 2000 Distance Experiments—Lau Technologies Identification Scores*

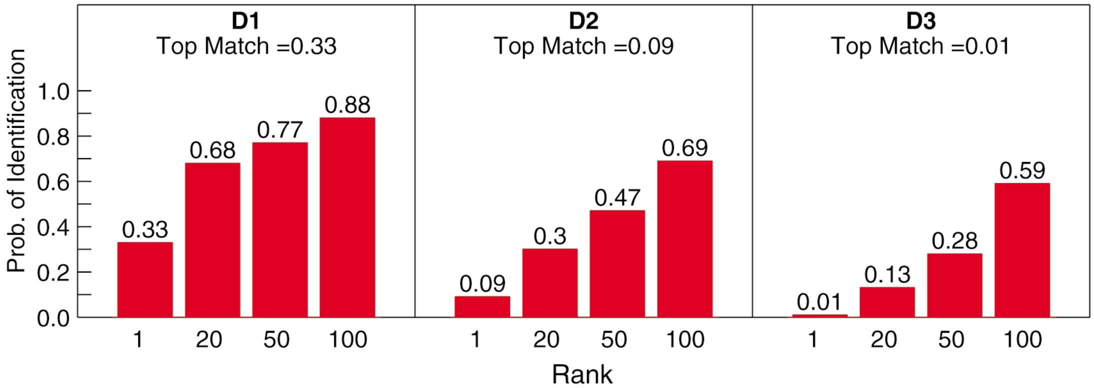

Figure 14: *FRVT 2000 Distance Experiments—Visionics Corp. Identification Scores*

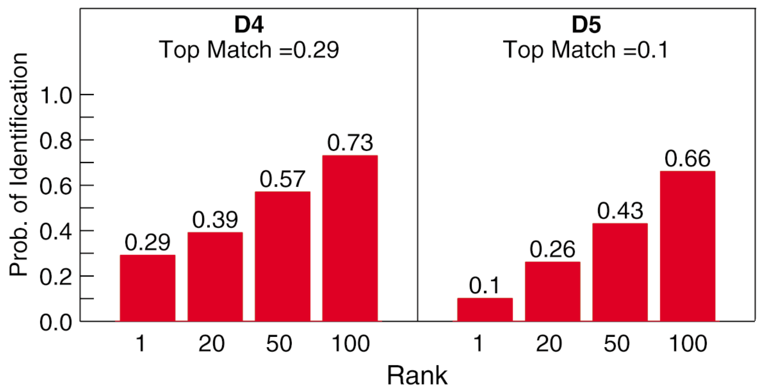

Figure 15: *FRVT 2000 Distance Experiments—Visionics Corp. Identification Scores*

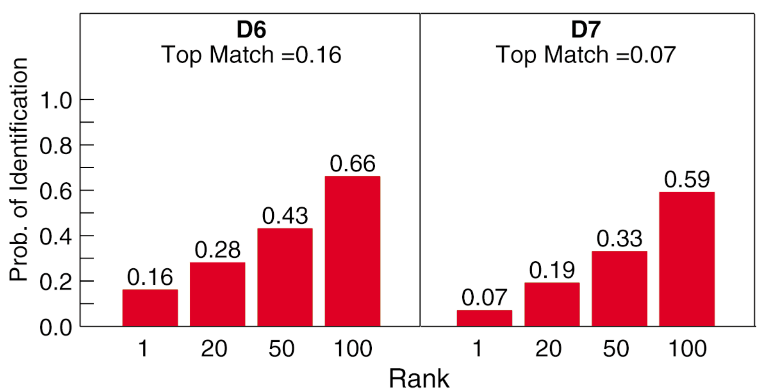

Figure 16: *FRVT 2000 Distance Experiments—Visionics Corp. Identification Scores*

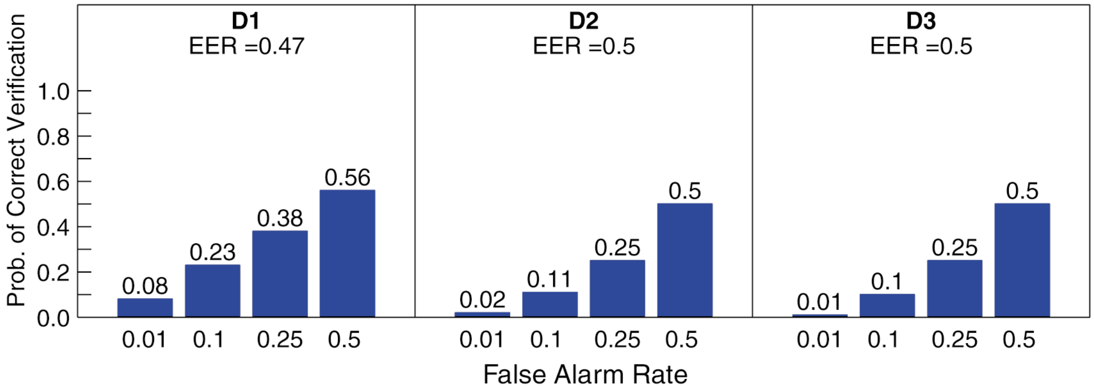

Figure 17: *FRVT 2000 Distance Experiments—C-VIS Verification Scores*

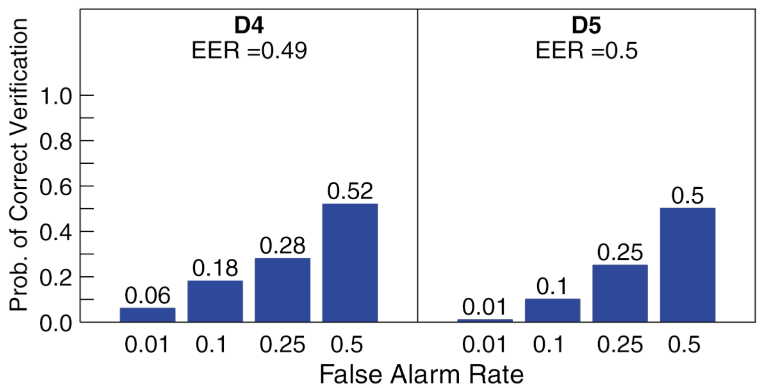

Figure 18: *FRVT 2000 Distance Experiments—C-VIS Verification Scores*

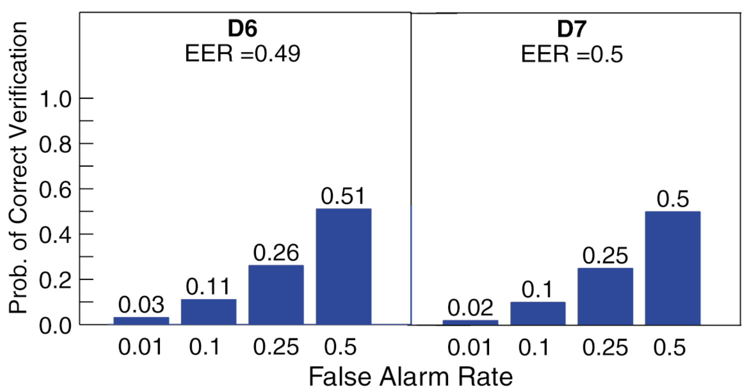

Figure 19: *FRVT 2000 Distance Experiments—C-VIS Verification Scores*

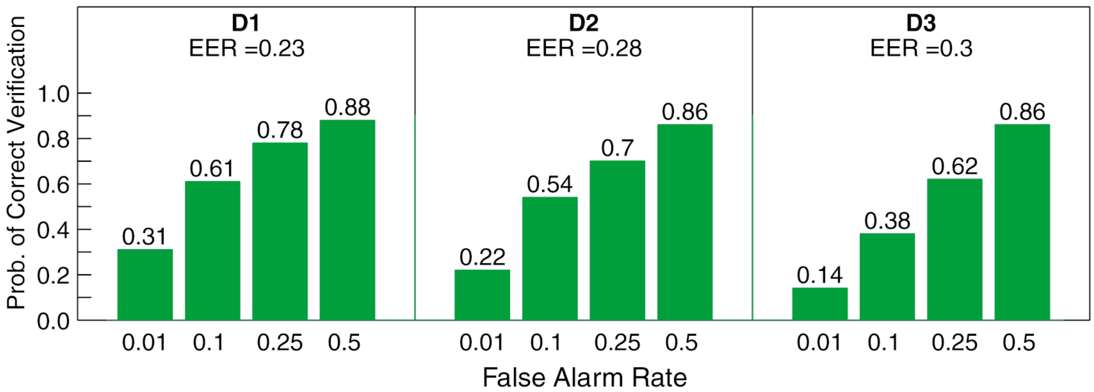

Figure 20: *FRVT 2000 Distance Experiments—Lau Technologies Verification Scores*

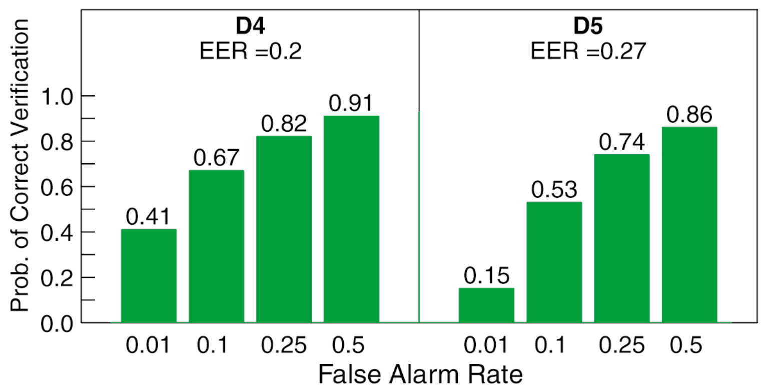

Figure 21: *FRVT 2000 Distance Experiments—Lau Technologies Verification Scores*

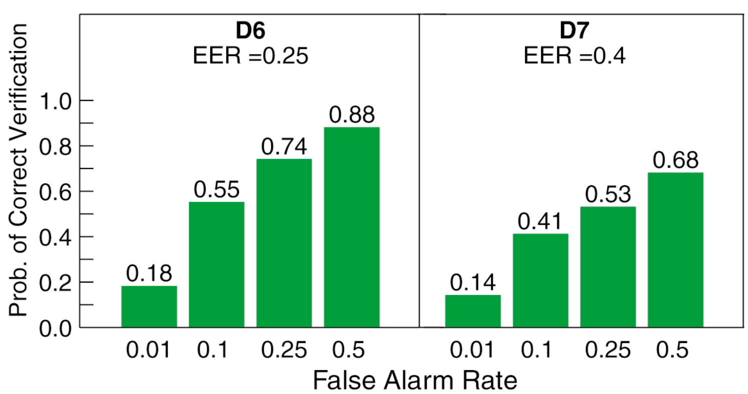

Figure 22: *FRVT 2000 Distance Experiments—Lau Technologies Verification Scores*

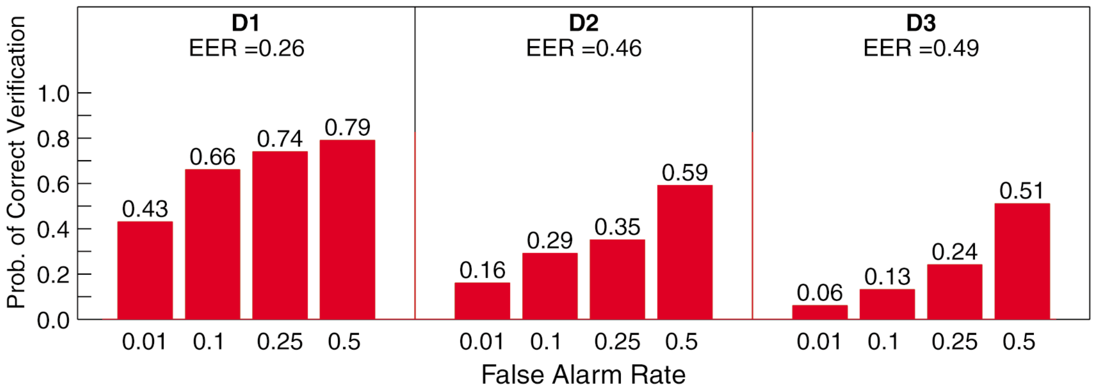

Figure 23: *FRVT 2000 Distance Experiments—Visionics Corp. Verification Scores*

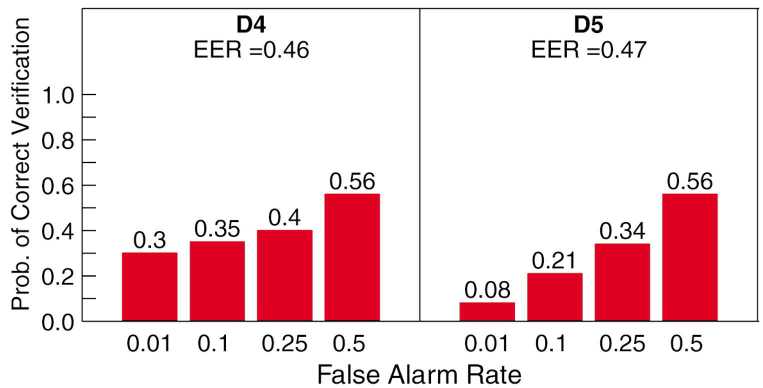

Figure 24: *FRVT 2000 Distance Experiments—Visionics Corp. Verification Scores*

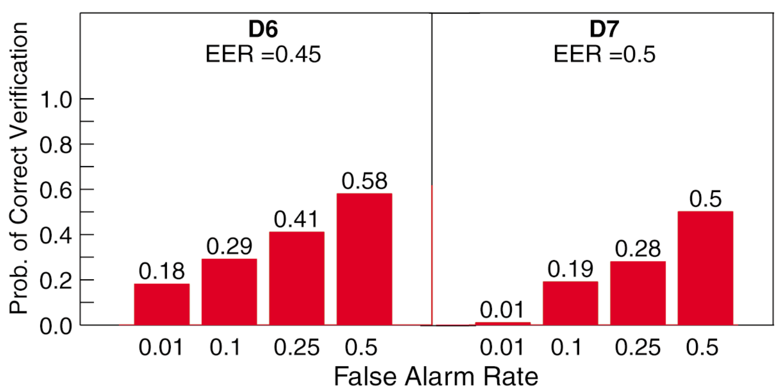

Figure 25: *FRVT 2000 Distance Experiments—Visionics Corp. Verification Scores*

7.1.3.3 Expression Experiments

The expression experiments were designed to evaluate the performance of face matching algorithms when comparing images of the same person with different facial expressions. This is an important consideration in almost any situation because it would be rare for a person to have the exact same expression for enrollment as for verification or identification.

The galleries and probe sets contain images of individuals captured at NIST in January 2000 and at Dahlgren in November 1999 using a digital CCD camera and two-lamp, FERET-style lighting. In this and other experiments, *fa* denotes a normal frontal facial expression, and *fb* denotes some other frontal expression.

Table 4: *Figures showing results of expression experiments. All images are frontal and were taken indoors with a digital camera using FERET-style lighting. The experiment consists of regular and alternate expressions (fa and fb images) from the same image set for each person.*

Experiment Name	Figure Numbers	Gallery Images	Probe Images	Gallery Size	Probe Set Size
E1	26, 27, 28, 29, 30, 31, M–19, M–41	Regular expression (fa image)	Alternate expression (fb image)	225	228
E2	26, 27, 28, 29, 30, 31, M–20, M–42	Alternate expression (fb image)	Regular expression (fa image)	224	228

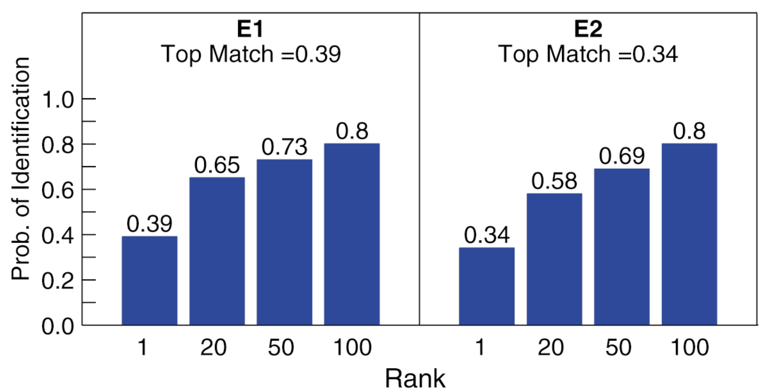

Figure 26: *FRVT 2000 Expression Experiments—C-VIS Identification Scores*

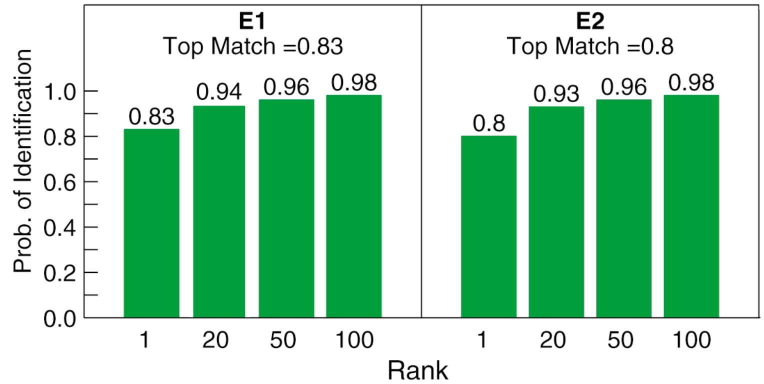

Figure 27: *FRVT 2000 Expression Experiments—Lau Technologies Identification Scores*

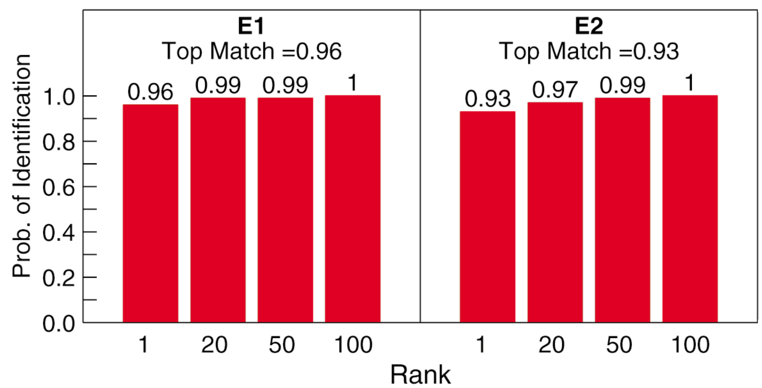

Figure 28: *FRVT 2000 Expression Experiments—Visionics Corp. Identification Scores*

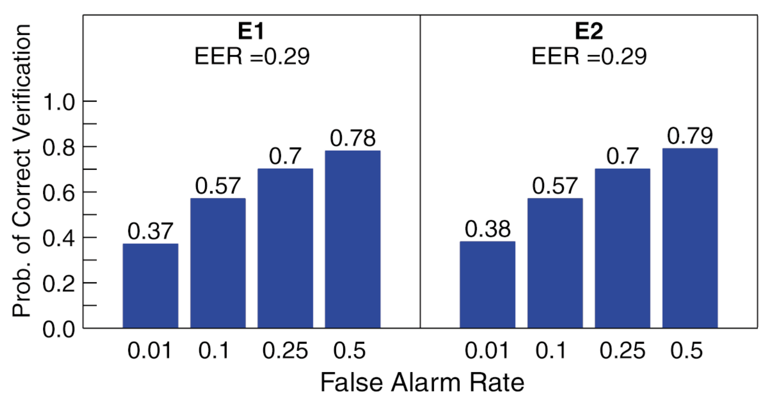

Figure 29: *FRVT 2000 Expression Experiments—C-VIS Verification Scores*

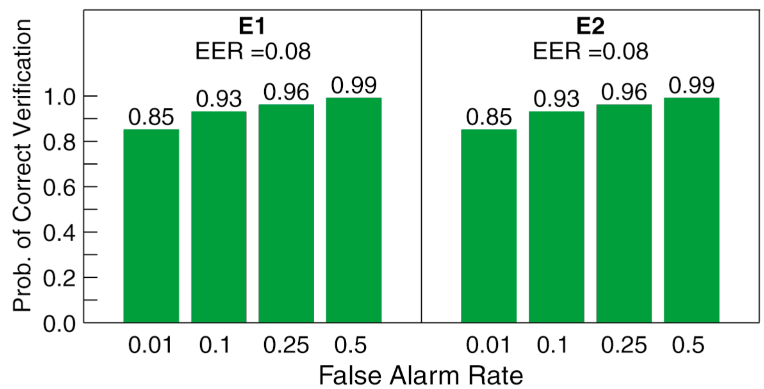

Figure 30: *FRVT 2000 Expression Experiments—Lau Technologies Verification Scores*

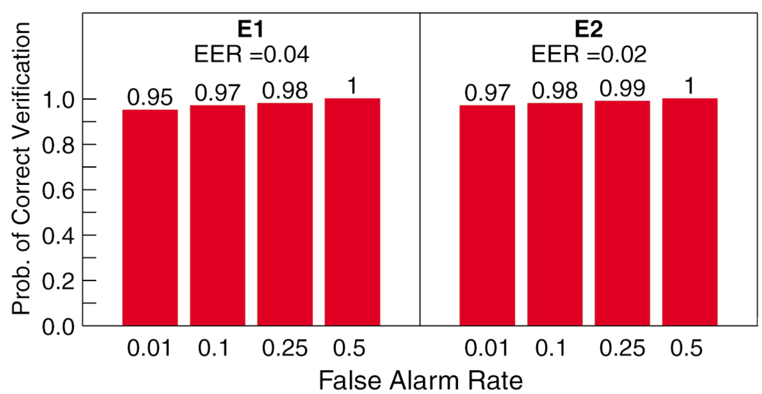

Figure 31: *FRVT 2000 Expression Experiments—Visionics Corp. Verification Scores*

7.1.3.4 Illumination Experiments

The problem of algorithm sensitivity to subject illumination is one of the most studied factors affecting recognition performance. When an image of the subject is taken under different lighting conditions than the condition used at enrollment, recognition performance can be expected to degrade. This is important for systems where the enrollment and the verification or identification are performed using different artificial lights, or when one operation is performed indoors and another outdoors.

The experiments described below use a single gallery containing high-quality, frontal digital stills of individuals taken indoors under mugshot lighting. The variation between experiments is through the probe sets, which are images taken shortly before or after their gallery matches using different lighting arrangements. In all cases, the individuals have normal facial expressions.

Table 5: *Figures showing results of illumination experiments. All images are frontal and were taken with a digital camera except when taken with the badging system.*

Experiment Name	Figure Numbers	Gallery Images	Probe Images	Gallery Size	Probe Set Size
I1	32, 33, 34, 35, 36, 37, M–21, M–43	Mugshot lighting	Overhead lighting	227	189
I2	32, 33, 34, 35, 36, 37, M–22, M–44	Mugshot lighting	Badge system lighting	129	130
I3	32, 33, 34, 35, 36, 37, M–23, M–45	Mugshot lighting	Outdoor lighting	227	190

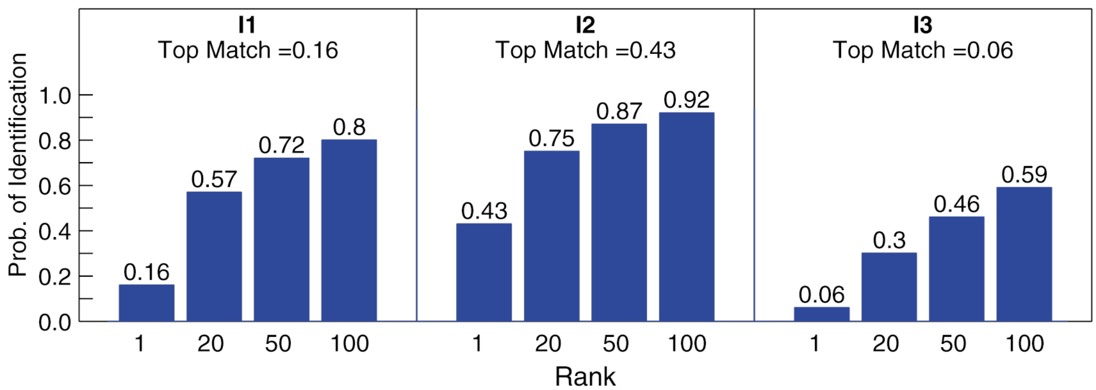

Figure 32: *FRVT 2000 Illumination Experiments—C-VIS Identification Scores*

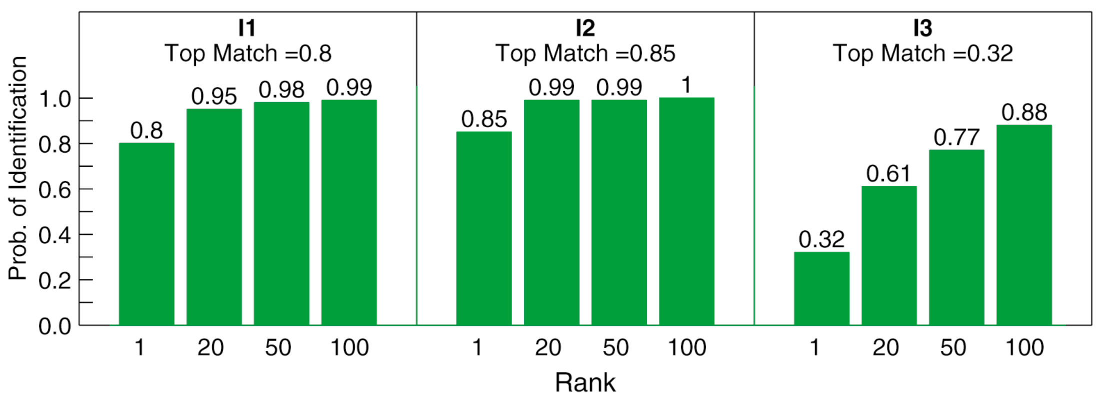

Figure 33: *FRVT 2000 Illumination Experiments—Lau Technologies Identification Scores*

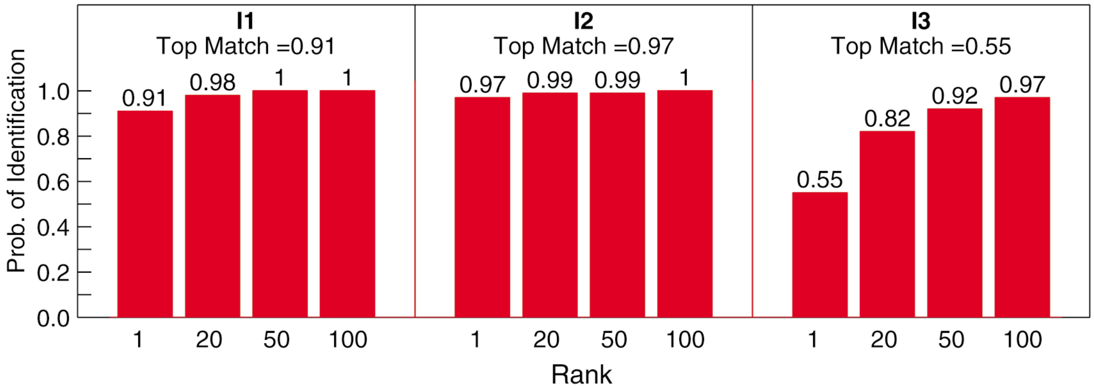

Figure 34: *FRVT 2000 Illumination Experiments—Visionics Corp. Identification Scores*

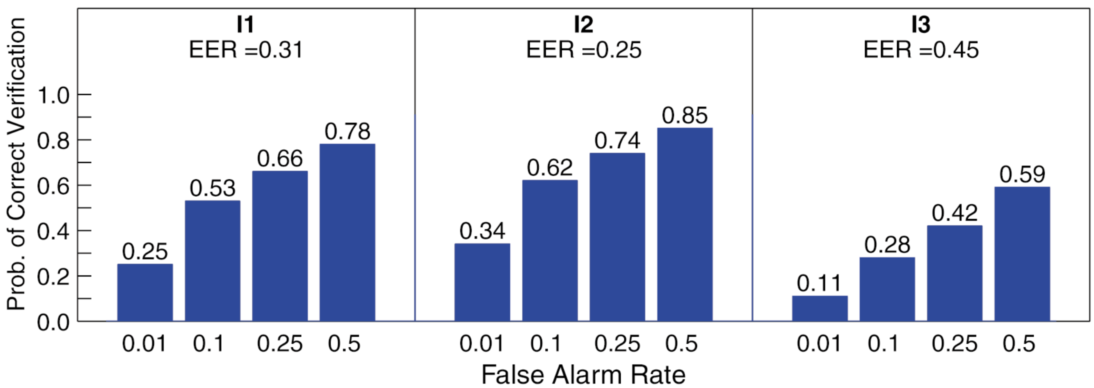

Figure 35: *FRVT 2000 Illumination Experiments—C-VIS Verification Scores*

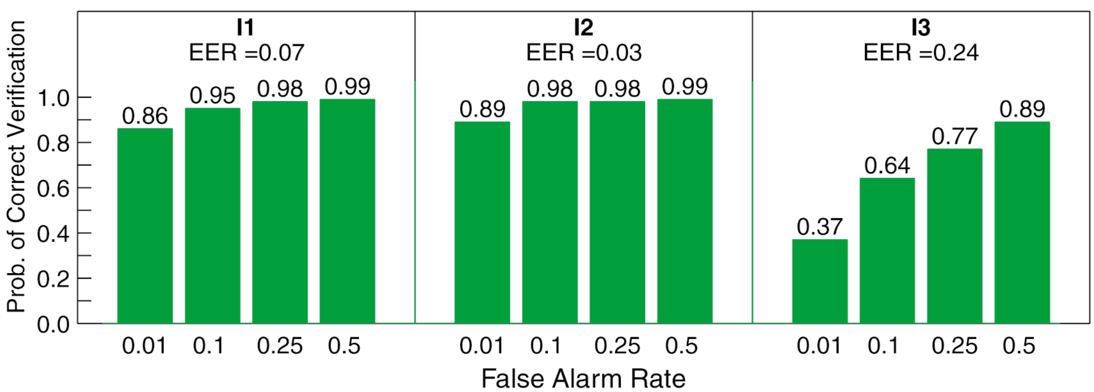

Figure 36: *FRVT 2000 Illumination Experiments—Lau Technologies Verification Scores*

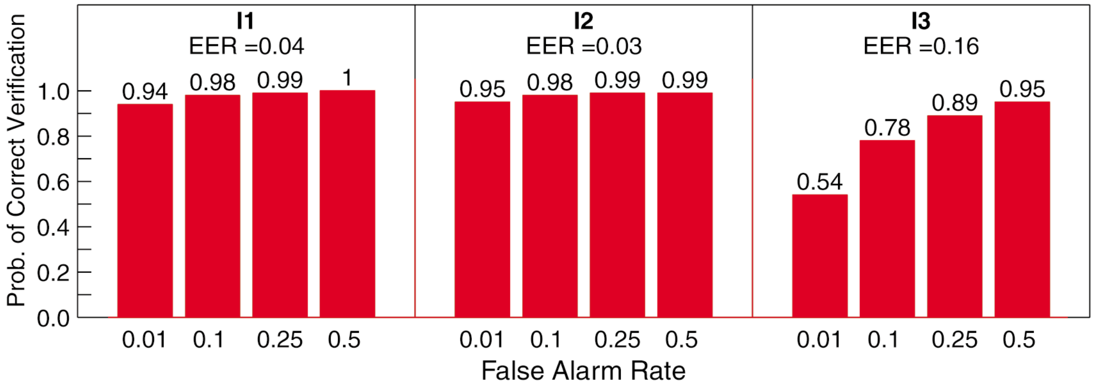

Figure 37: *FRVT 2000 Illumination Experiments—Visionics Corp. Verification Scores*

7.1.3.5 Media Experiments

The media experiments were designed to evaluate the performance of face-matching algorithms when comparing images stored on different media. In this case, digital CCD images and 35mm film images are used. This is an important consideration for a scenario such as using an image captured with a video camera to search through a mugshot database created from a film source.

The galleries for the media experiments are made up of images taken at Dahlgren in November 1999 and NIST in December 2000 of individuals wearing normal (*fa*) facial expressions indoors. The galleries contain either film images or digital CCD images; the probe contains the other. Usually the images were taken simultaneously within a few tenths of a second of each other.

Table 6: *Figures showing results of media experiments. All images were taken indoors and are frontal regular expression (fa) images. All images of a person are from the same set. The gallery and probe camera columns show the camera type used to acquire the images.*

Experiment Name	Figure Numbers	Gallery Camera	Probe Camera	Gallery Size	Probe Set Size
M1	38, 39, 40, 41, 42, 43, M–24, M–46	35mm	Digital	96	102
M2	38, 39, 40, 41, 42, 43, M–25, M–47	Digital	35mm	227	99

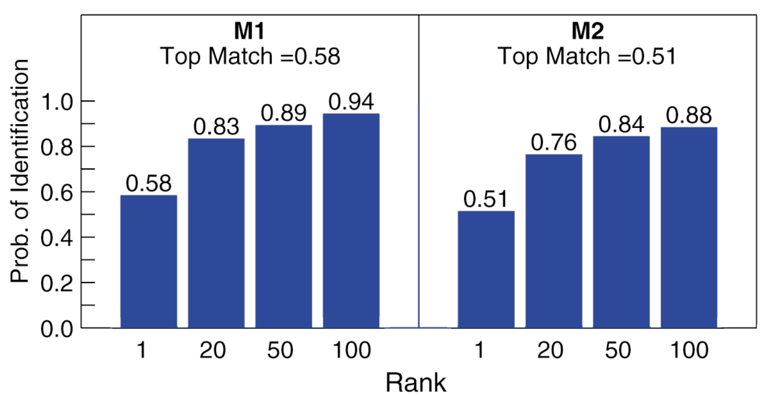

Figure 38: *FRVT 2000 Media Experiments—C-VIS Identification Scores*

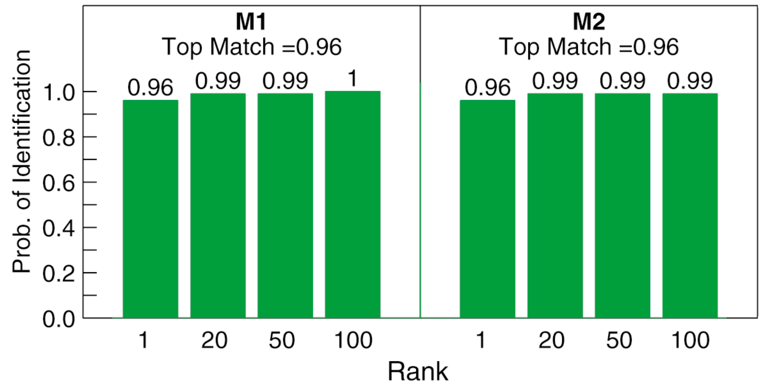

Figure 39: *FRVT 2000 Media Experiments—Lau Technologies Identification Scores*

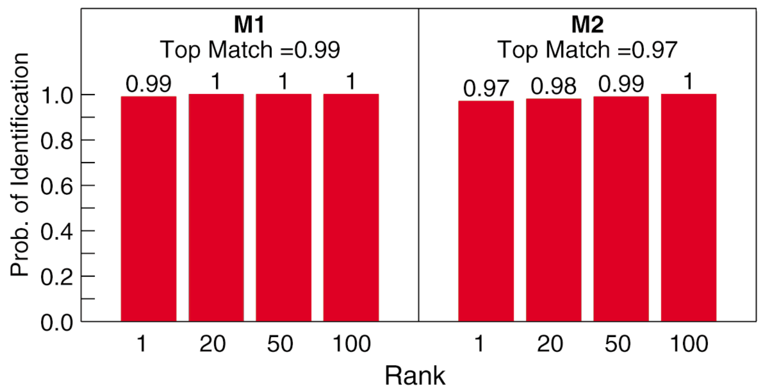

Figure 40: *FRVT 2000 Media Experiments—Visionics Corp. Identification Scores*

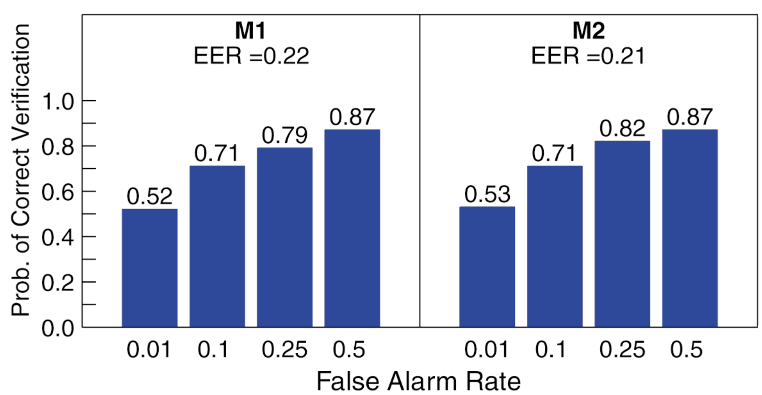

Figure 41: *FRVT 2000 Media Experiments—C-VIS Verification Scores*

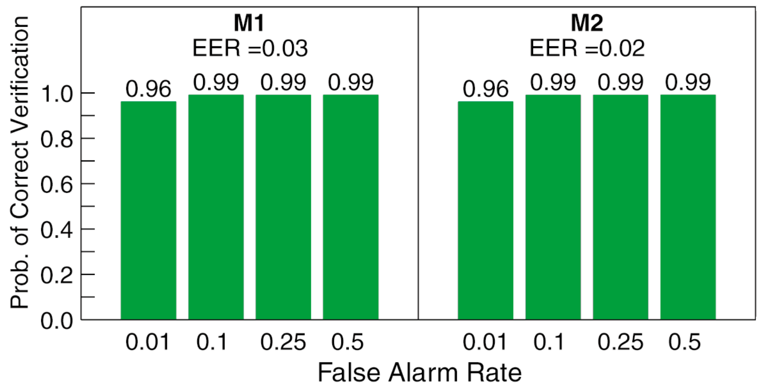

Figure 42: *FRVT 2000 Media Experiments—Lau Technologies Verification Scores*

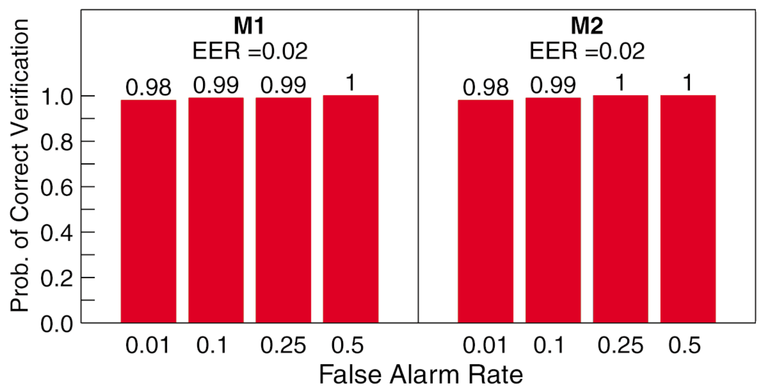

Figure 43: *FRVT 2000 Media Experiments—Visionics Corp. Verification Scores*

7.1.3.6 Pose Experiments

The performance of face-matching algorithms applied to images of subjects taken from different viewpoints is of great interest in certain applications, most notably those using indifferent or uncooperative subjects, such as surveillance. Although a subject may look up or down and thereby vary the declination angle, the more frequently occurring and important case is where the subject is looking ahead but is not facing the camera. This variation is quantified by the azimuthal head angle, referred to here as the pose. The experiments described below address the effect of pose variation. These experiments do not address angle of declination or a third variation—side-to-side head tilt.

The imagery used in the pose experiments were taken from two sources. For studies P1-P4, the b15 subset of the FERET collection was used. These images were obtained from 200 individuals who were asked to face in nine different directions under tightly controlled conditions. The P1-P4 gallery contains only frontal images. Each probe set contains images from one of the four different, non-frontal orientations. No distinction was made between left- and right-facing subjects on the assumption that many algorithms behave symmetrically.

The P5 study is distinct because its imagery is not from the FERET collection. Its gallery holds frontal outdoor images, while the probe set contains a corresponding image of the subject facing left or right at about 45 degrees to the camera.

Table 7: *Figures showing results of pose experiments. All images of a person are from the same image set. The image-type colum refers to gallery and probe images. FERET refers to the FERET database and HumanID the HumanID database (new images included in the FRVT 2000). Pose angles are in degrees with 0 being a frontal image.*

Experiment Name	Figure Numbers	Image Type	Gallery Pose	Probe Pose	Gallery Size	Probe Set Size
P1	44, M–6	FERET	0	15	200	400
P2	44, M–7	FERET	0	25	200	400
P3	44, M–8	FERET	0	40	200	400
P4	44, M–9	FERET	0	60	200	400
P5	45, 46, 47, 48, 49, 50, M–26, M–48	HumanID, digital, outdoors	0	45	180	186

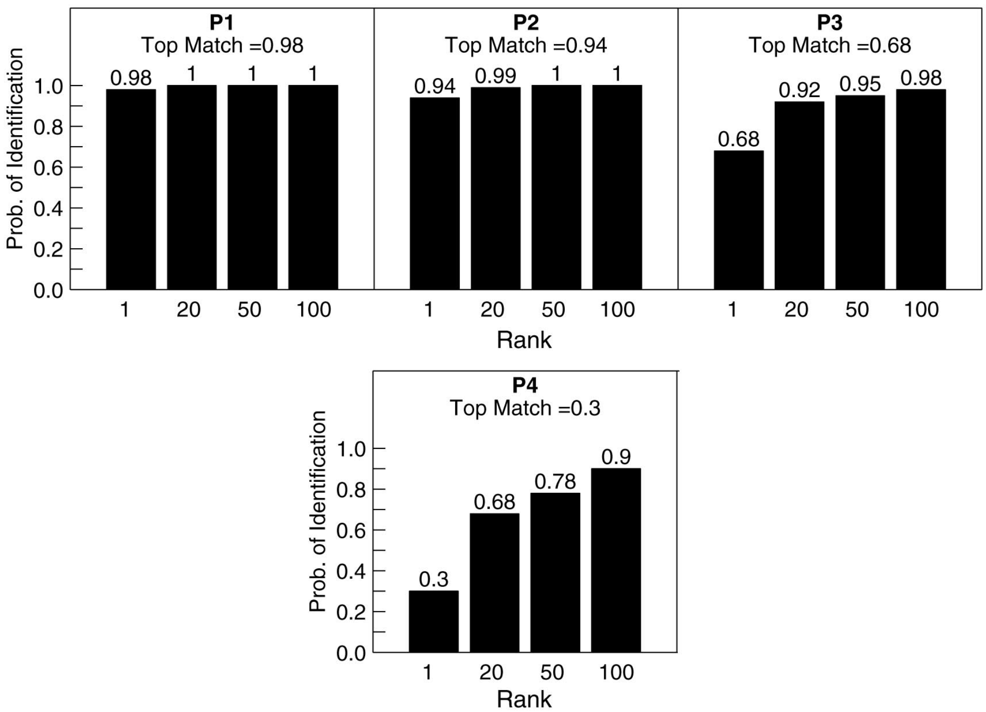

Figure 44: *FERET Results—Pose Experiments Best Identification Scores*

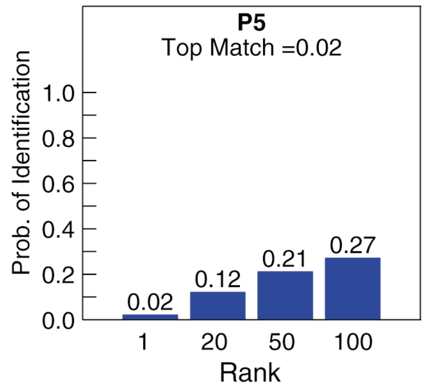

Figure 45: *FRVT 2000 Pose Experiments—C-VIS Identification Scores*

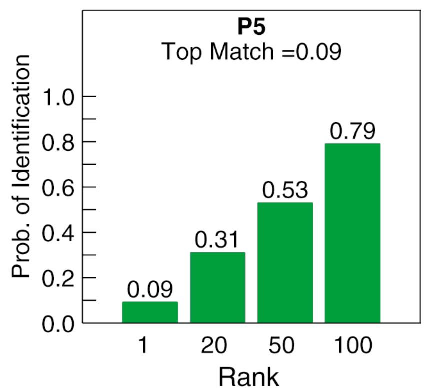

Figure 46: *FRVT 2000 Pose Experiments—Lau Technologies Identification Scores*

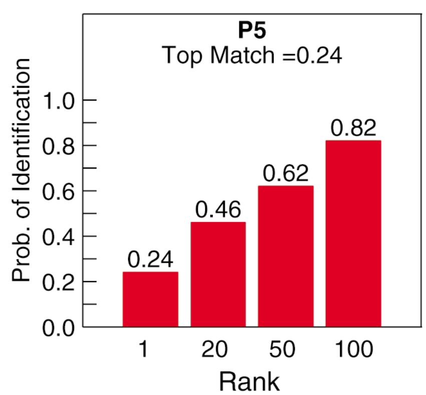

Figure 47: *FRVT 2000 Pose Experiments—Visionics Corp. Identification Scores*

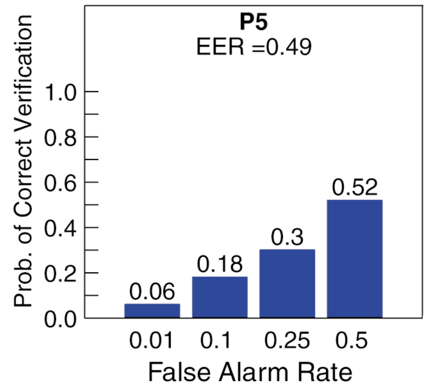

Figure 48: *FRVT 2000 Pose Experiments—C-VIS Verification Scores*

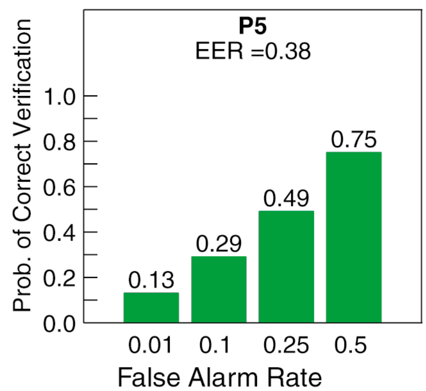

Figure 49: *FRVT 2000 Pose Experiments—Lau Technologies Verification Scores*

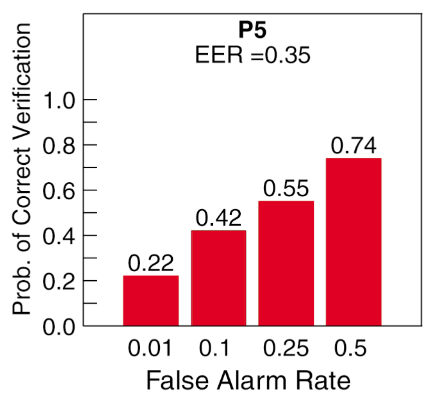

Figure 50: *FRVT 2000 Pose Experiments—Visionics Corp. Verification Scores*

7.1.3.7 Resolution Experiments

Image resolution is critical to face recognition systems. There is always some low resolution at which the face image will be of sufficiently small size that the face is unrecognizable. The resolution experiments described below were designed to evaluate the performance of face matching as resolution is decreased. The metric we have used to quantify resolution is eye-to-eye distance in pixels. The imagery used is homogenous in the sense that it was all taken at a fixed distance to a camera, and the resolution is decreased off-line using a standard reduction algorithm. This procedure is driven by the manually keyed pupil coordinates present in the original imagery. The fractional reduction in size is determined simply as the ratio of the original and sought eye-to-eye distances. The resulting eye-to-eye distances are as low as 15 pixels.

A single, high-resolution gallery is used for all the resolution tests. It contains full-resolution, digital CCD images taken indoors under mugshot standard flood lighting. The gallery eye separation varies according to the subject with a mean of 138.7 pixels and a range of 88 to 163. In all cases, the probe sets are derived from those same gallery images. The aspect ratio is preserved in the reduction. Note that subjects with large faces are reduced by a greater factor than those with small heads.

Table 8: *Figures showing results of resolution experiments. All images of a person are from the same set. The distance between the centers of the eyes in the rescaled probes is expressed in pixels in the probe eye separation column.*

Experiment Name	Figure Numbers	Probe Eye Separation	Gallery Size	Probe Set Size
R1	51, 52, 53, 54, 55, 56, M–27, M–49	60	101	102
R2	51, 52, 53, 54, 55, 56, M–28, M–50	45	101	102
R3	51, 52, 53, 54, 55, 56, M–29, M–51	30	101	102
R4	51, 52, 53, 54, 55, 56, M–30, M–52	15	101	102

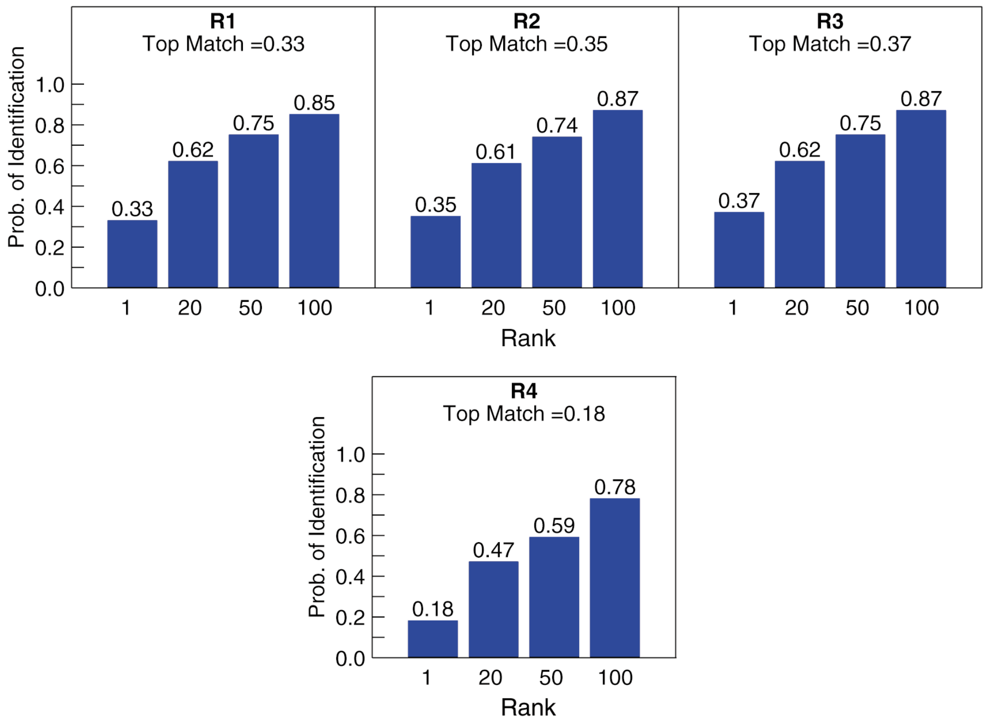

Figure 51: *FRVT 2000 Resolution Experiments—C-VIS Identification Scores*

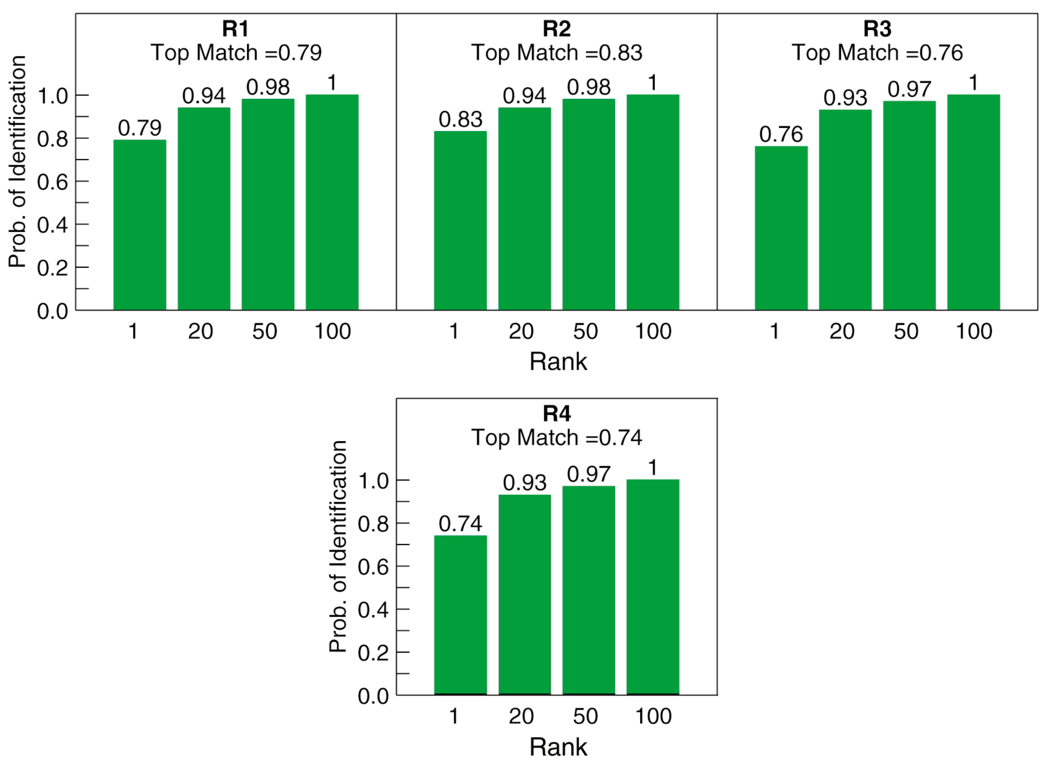

Figure 52: *FRVT 2000 Resolution Experiments—Lau Technologies Identification Scores*

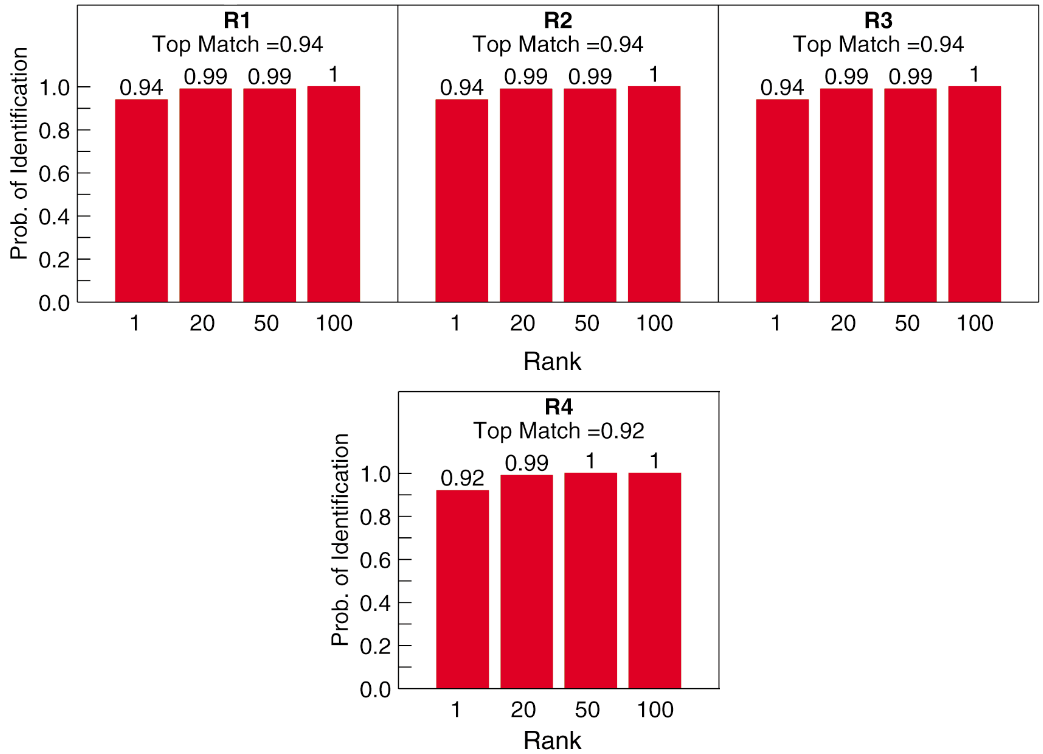

Figure 53: *FRVT 2000 Resolution Experiments—Visionics Corp. Identification Scores*

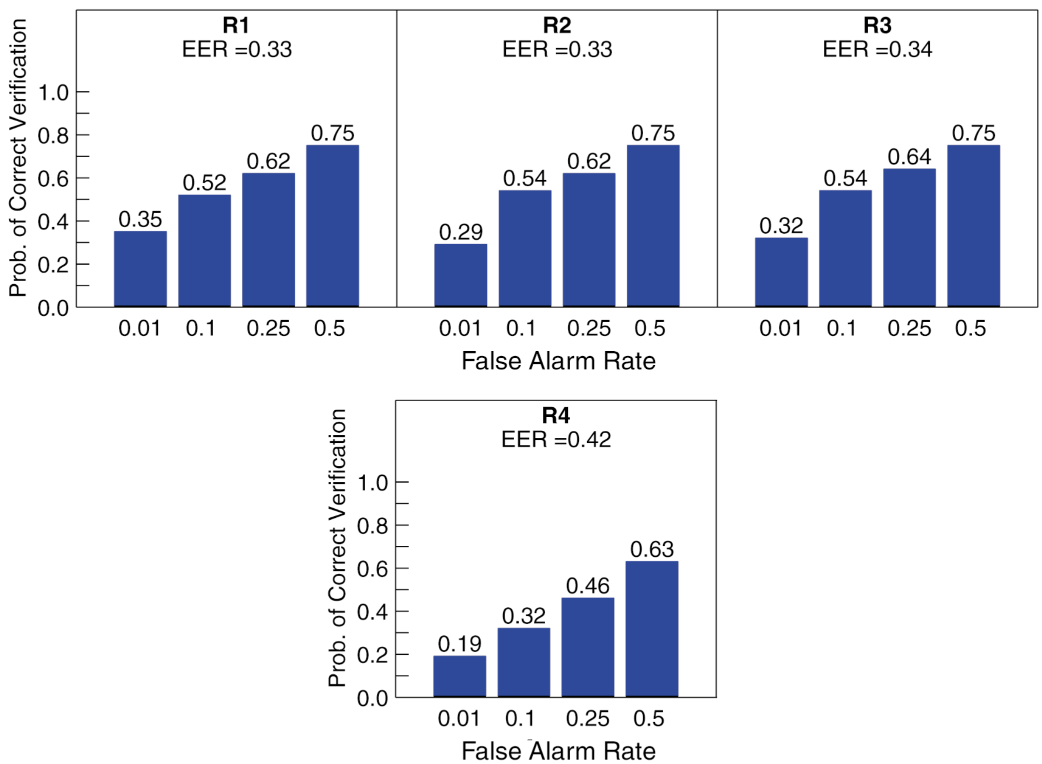

Figure 54: *FRVT 2000 Resolution Experiments—C-VIS Verification Scores*

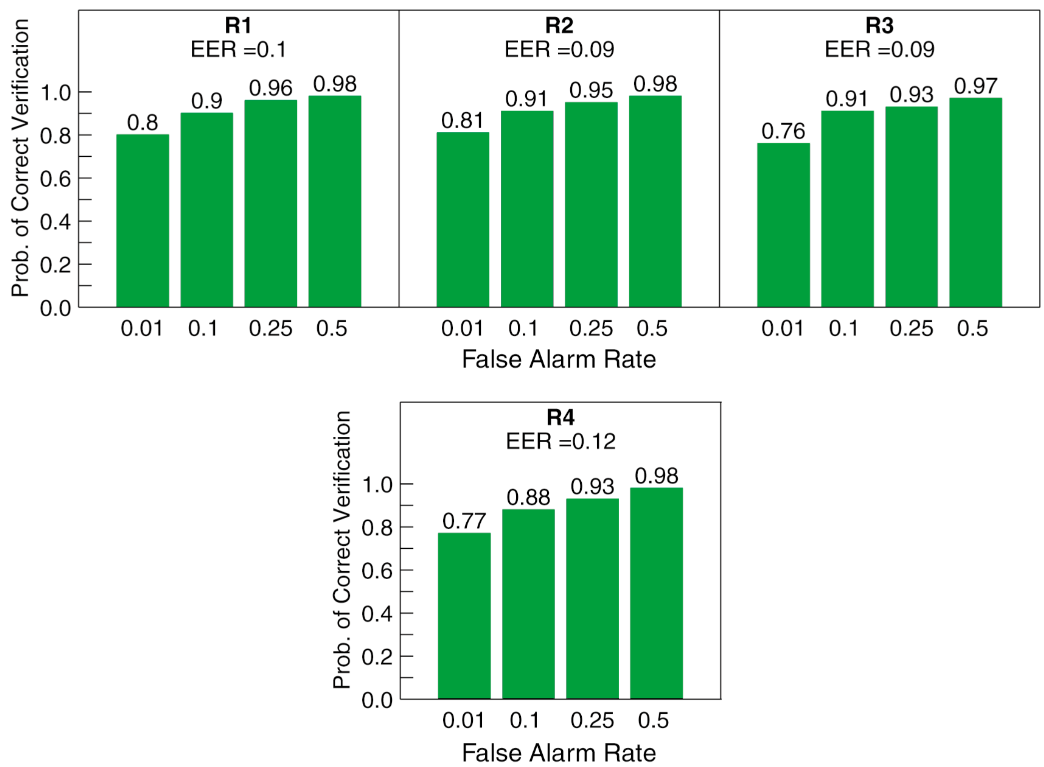

Figure 55: *FRVT 2000 Resolution Experiments—Lau Technologies Verification Scores*

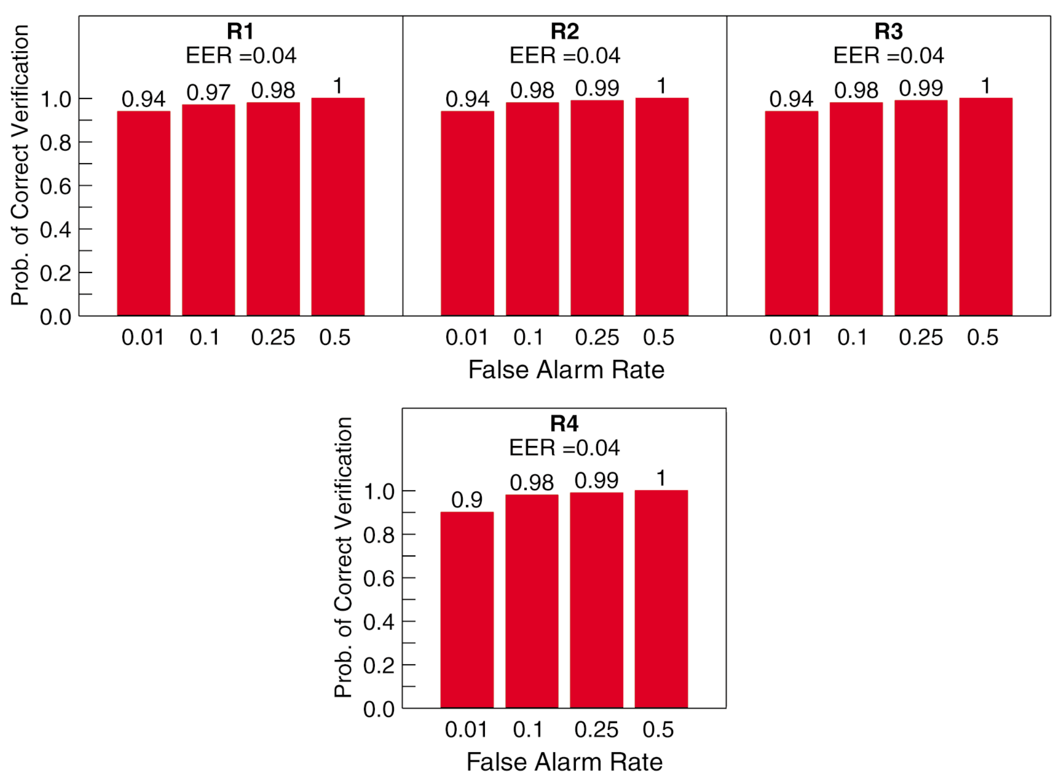

Figure 56: *FRVT 2000 Resolution Experiments—Visionics Corp. Verification Scores*

7.1.3.8 Temporal Experiments

The temporal experiments address the effect of time delay between first and subsequent captures of facial images. The problem of recognizing subjects during extended periods is intuitively significant and is germane to many applications. Robust testing of this effect is difficult because of a lack of long-term data. Given the absence of meaningful data sets, these experiments rely on imagery gathered during a period of less than two years.

The T1 and T2 studies exactly reproduce the widely reported FERET duplicate I and II tests. They use the standard frontal 1,196-image FERET gallery.

The T2 probe set contains 234 images from subjects whose gallery match was taken between 540 and 1,031 days before (median = 569, mean = 627 days). The T1 probe set is a superset of the T2 probe set with additional images taken closer in time to their gallery matches. The T1 probe set holds 722 images whose matches were taken between 0 and 1031 days after the match (median = 72, mean = 251 days). The difference set (T1-T2 has 488 images) has time delays between 0 and 445 days (median = 4, mean = 70 days). Thus T2 is a set where at least 18 months has elapsed between capturing the gallery match and the probe itself. T1 and T2 also represent an access control situation in which a gallery is rebuilt every year or so.

Experiments T3-T5 are based on the more recent HumanID image collections. The galleries contain about 227 images that were obtained between 11 and 13 months after the probe images. The probe set is fixed and contains 467 images obtained using overhead room lighting. The three studies differ only in the lighting used for the gallery images.

Table 9a: *Figures showing results of temporal experiments.*

Experiment Name	Figure Numbers	Experiment Description	Gallery Size	Probe Set Size
T1	57, M–10	FERET Duplicate I	1,196	722
T2	57, M–11	FERET Duplicate II	1,196	234

Table 9b: *Figures showing results of temporal experiments. The T3-T5 experiment gallery was made up of digital frontal images collected at Dahlgren in 1999 and NIST in 2000. The probe images are frontal and were collected at Dahlgren in 1998.*

Experiment Name	Figure Numbers	Gallery Lighting	Probe Lighting	Gallery Size	Probe Set Size
T3	58, 59, 60, 61, 62, 63, M–31, M–53	Mugshot	Ambient	227	467
T4	58, 59, 60, 61, 62, 63, M–32, M–54	FERET	Ambient	227	467
T5	58, 59, 60, 61, 62, 63, M–33, M–55	Overhead	Ambient	226	467

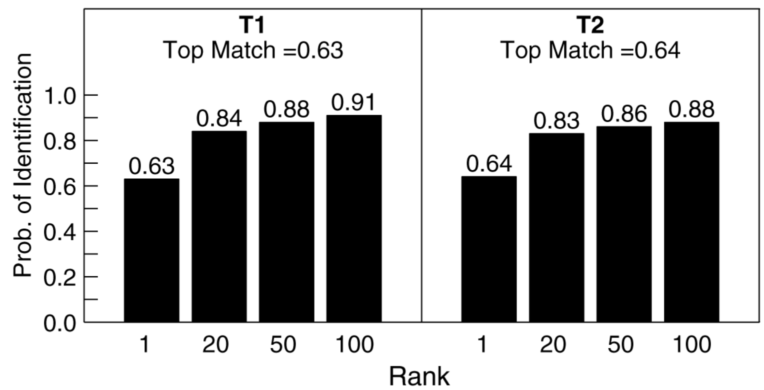

Figure 57: *FERET Results—Temporal Experiments Best Identification Scores*

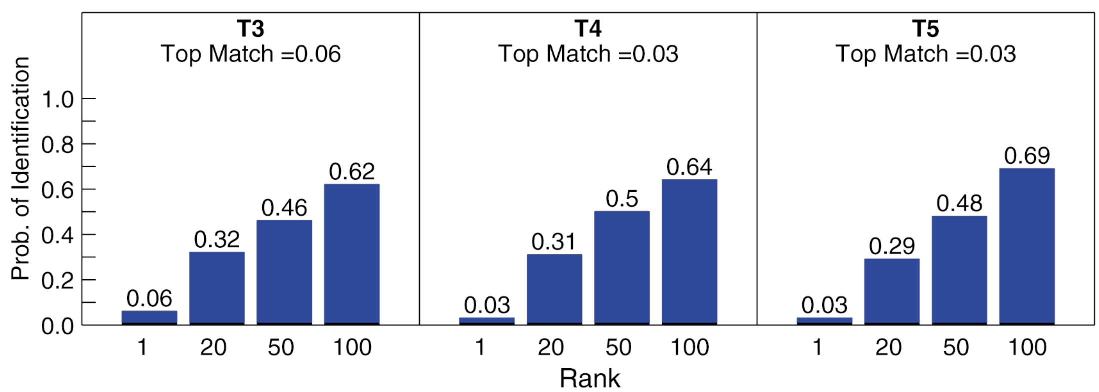

Figure 58: *FRVT 2000 Temporal experiments—C-VIS Identification Scores*

Figure 59: *FRVT 2000 Temporal experiments—Lau Technologies Identification Scores*

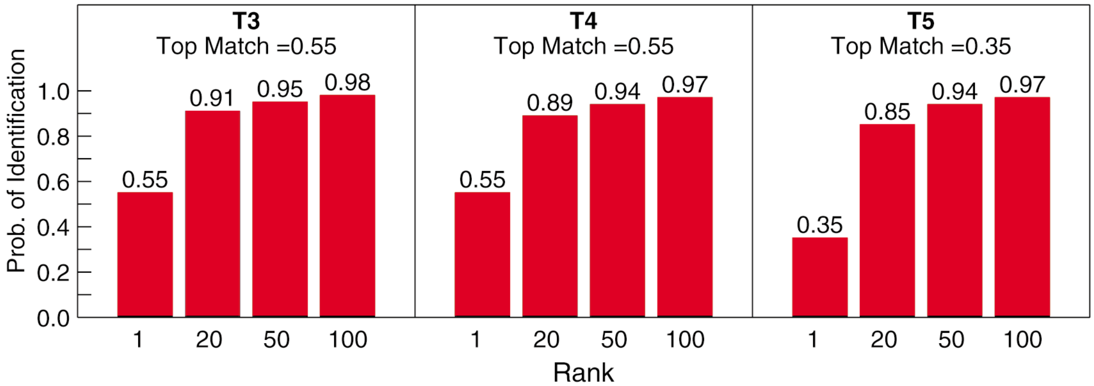

Figure 60: *FRVT 2000 Temporal experiments—Visionics Corp. Identification Scores*

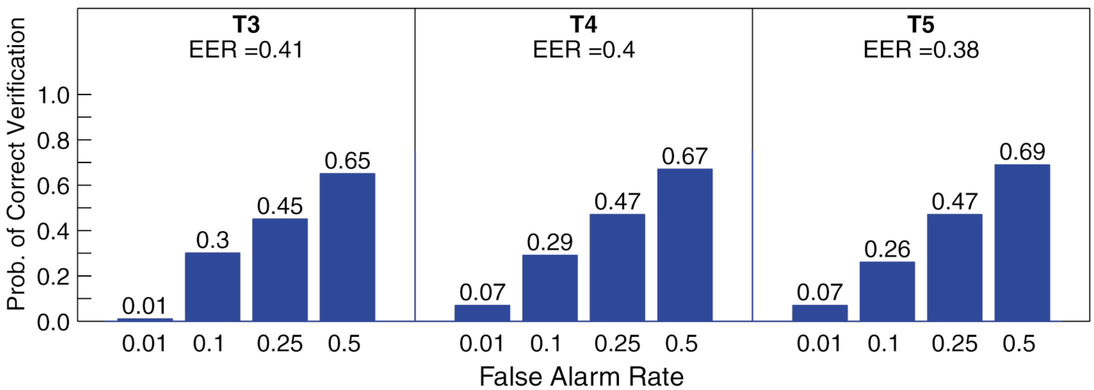

Figure 61: *FRVT 2000 Temporal experiments—C-VIS Verification Scores*

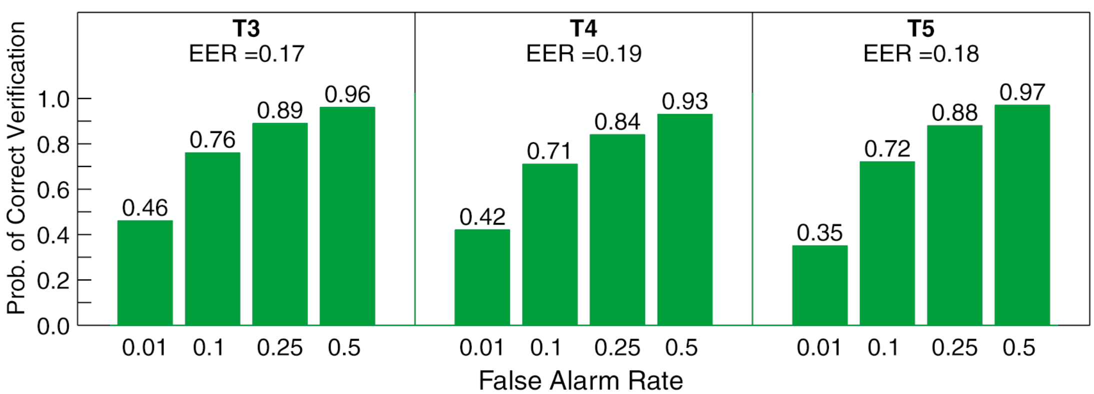

Figure 62: *FRVT 2000 Temporal experiments—Lau Technologies Verification Scores*

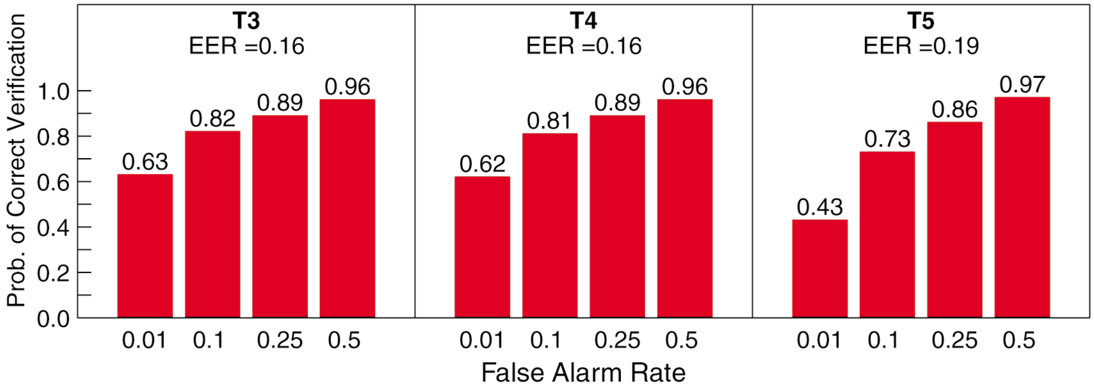

Figure 63: *FRVT 2000 Temporal experiments—Visionics Corp. Verification Scores*

7.2 Product Usability Test

7.2.1 Overview

The scenario chosen for the Product Usability Test was access control with live subjects. Some systems tested, however, were not intended for access control applications. The intended application for each system, as shown in Appendix J, should be kept in mind when evaluating the results of the Product Usability Test.

The Product Usability Test was administered in two parts: the Old Image Database Timed Test and the Enrollment Timed Test. For the Old Image Database Timed Test, vendors were given a set of 165 images captured with a standard access control badge system, including one image of each of the three test subjects. The set contained two images for five people, and one image for each of the other 155 people. Vendors enrolled these images into their system for comparison with the live subjects. The operational scenario was that of a low-security access control point into the lobby of a building. The building's security officers did not want to mandate that the employees take the time to enroll into the new facial recognition system so they used their existing digital image database taken from the employee's picture ID badges.

For the Enrollment Timed Test, the images of the three test subjects were removed from the system while the other images were retained. Vendors were then allowed to enroll the three subjects using their standard procedures, including the use of multiple images. The purpose of the test was to measure system performance using vendor enrollment procedures. The enrollment procedures were not evaluated. The operational scenario was that of an access control door for a medium-to-high security area within the building previously described. In this case, employees were enrolled in the facial recognition system using the standard procedures recommended by the vendor.

During the Product Usability Test, several parameters were varied including start distance, behavior mode, and backlighting. Tests were performed for each subject at distances of 12, 8, and 4 feet for all trials except for the variability test. Test subjects performed each test—always at 12 feet—using cooperative and simulated, repeatable, indifferent behavior modes. For the cooperative mode, subjects looked directly at the camera for the duration of the trial. For the indifferent mode (we will refer to this as indifferent from this point forward), subjects instead moved their focus along a triangular path made up of three visual targets surrounding the camera. Each trial was performed with and without backlighting provided by a custom light box.

For the Old Image Database Timed Test, subjects began each trial standing at the specified start distance then walked toward the camera when the timer was started. Each subject started at 12, 8 and 4 feet in cooperative mode then repeated in indifferent mode. Subject 1 then performed 8 cooperative trials from a start distance of 12 feet for the variability test, a test to determine the consistency of the subject-system interaction. Subject 1 then performed three more cooperative trials from 12, 8, and 4 feet holding a photograph of his own face to determine if the system could detect liveness. The photograph was an 8" x 10" color glossy print taken in a professional photo studio. This entire sequence was followed four times: once in verification mode without backlighting, once in identification mode without backlighting, once in verification mode with backlighting, and once in identification mode with backlighting.

The Enrollment Timed Test was performed exactly as the Old Image Database Timed Test described above except the subjects stood in place at the specified start distance rather than walking toward the camera.

7.2.2 Interpreting the Results – What Do the Tables Mean?

The tables in Section 7.2.4 and Section 7.2.5 show the data recorded during the live tests. For the Old Image Database Timed Test, three parameters were recorded:

- *Final distance* is the distance in feet between the camera and the test subject at the end of the trial. This was recorded in increments of one foot.
- *Acquire time* is the time in seconds it took the system to report a match, regardless of whether or not the answer was correct. This was recorded in increments of 1/100 second. An X indicates that a match was not acquired within the 10-second time limit.
- *Correct match* tells whether or not the system matched the live subject with the correct person in the database. Again, an X indicates that a match was not acquired within the 10 second time limit.

For the Enrollment Timed Test, the parameters were recorded as described; however, the subjects stood in place for each of these trials so it was unnecessary to record the final distance.

For the variability test, subject 1 performed eight cooperative-mode trials for both the verification and identification modes, with and without backlighting. A start distance of 12 feet was used for each trial.

Note that it is desirable to have a correct match on all trials except the photo tests, where a photo of subject 1 was used to attempt access. Although none of the vendors claimed to have a liveness detection feature, most systems were not fooled by the photo.

Also note that most systems performed much better in the Enrollment Timed Test than in the Old Image Database Timed Test. This is most likely because the Old Image Database Timed Test used a database with one image per subject taken with a different camera and under different lighting conditions than those used in the testing room. For the Enrollment Timed Test, subjects were enrolled and tested for a match in the same testing room and multiple images were taken in most cases.

7.2.3 Sample Images and Test Subject Description

For the Old Image Database Timed Test, vendors were given a set of 165 images of 160 people (including one image of each of the three test subjects) to use for enrollment. These images were

acquired using a standard access-control badge system developed and maintained by NAVSEA Crane. The system is made up of the following components:

- EBACS Mk3 Mod 4 badge software (developed by NAVSEA Crane);
- Integral Technologies' FlashPoint 3075 PCI video frame grabber;
- Imaging Technology Corp.'s CCD 1000 video camera;
- Lowel iLIGHT portrait lighting system, including a single 100W, 3200K lamp.

In each case, images were collected at two different sites using the same system, with overhead fluorescent lighting in addition to the system lamp. There were 33 images of 33 subjects acquired at NAVSEA Crane, and 132 images of 127 subjects acquired at NIST. One image per subject was acquired at NAVSEA Crane. One image was acquired for each of 122 subjects at NIST, while two images were acquired for five subjects. Subjects stood 8 feet in front of a camera adjusted to a height of 5 ft. 6 in. A white wall was located one foot behind the subject. Images were captured with a resolution of 380 x 425 and saved as 24-bit JPEG files with a quality setting of 90 percent.

Figure 64 shows the color images of the three test subjects used for the Old Image Database Timed Test. Subject 1 is a 6-ft. Caucasian male with glasses. Subject 2 is a 6 ft.-1 in. Caucasian male without glasses. Subject 3 is a 5 ft.-2 in. Caucasian female without glasses.

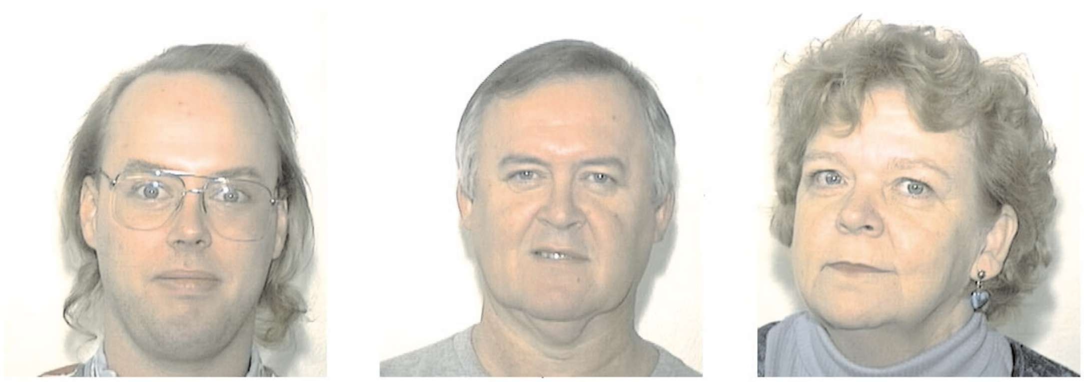

Figure 64: *Sample Images from EBACS Mk3 Mod 4 badging system. From left to right, subject 1, subject 2 and subject 3.*

7.2.4 Old Image Database Timed Test Results

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	1	7.53	No	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	10.00	No
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	10.00	No
2	Cooperative	12	1	6.41	No	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	9.95	No
	Indifferent	12	1	X	X	1	X	X
		8	1	5.19	No	1	X	X
		4	1	X	X	1	X	X
3	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	3.83	No	1	X	X
		4	1	X	X	1	X	X
1 Variability Test	Cooperative	12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	4	4.87	No	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
1 Photo Test	Cooperative	12	1	8.37	Yes	1	9.55	Yes
		8	3	3.77	Yes	1	X	X
		4	1	X	X	1	3.35	Yes

Table 10: *Banque-Tec—Old Image Database Timed Test Verification Mode*

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	1	10.00	No	3	6.62	No
		8	1	6.63	No	1	5.90	No
		4	1	7.69	No	1	5.44	No
	Indifferent	12	1	X	X	3	7.45	No
		8	1	X	X	1	8.26	No
		4	1	6.73	No	1	6.12	No
2	Cooperative	12	1	X	X	7	4.09	No
		8	1	8.76	No	1	8.33	No
		4	1	5.09	No	1	5.18	No
	Indifferent	12	1	8.26	No	3	7.04	No
		8	1	9.46	No	1	9.13	No
		4	1	8.14	No	1	6.48	No
3	Cooperative	12	1	9.33	No	1	10.00	No
		8	1	8.13	No	1	5.44	No
		4	1	5.58	No	1	5.09	No
	Indifferent	12	1	X	X	2	8.00	No
		8	1	10.00	No	1	7.15	No
		4	1	7.37	No	1	7.58	No
1 Variability Test	Cooperative	12	3	5.66	No	4	5.51	No
		12	2	8.00	No	3	5.54	No
		12	4	5.95	No	3	5.42	No
		12	4	4.92	No	5	3.95	No
		12	3	6.56	No	1	8.28	No
		12	3	6.58	No	3	5.78	No
		12	2	6.53	No	4	5.47	No
		12	3	6.19	No	1	9.75	No
1 Photo Test	Cooperative	12	6	4.09	No	7	3.74	No
		8	2	5.50	No	4	3.80	No
		4	1	4.02	No	1	5.58	No

Table 11: *C-VIS—Old Image Database Timed Test Verification Mode*

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
2	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
3	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
1 Variability Test	Cooperative	12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
1 Photo Test	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X

Table 12: *Lau Technologies—Old Image Database Timed Test Verification Mode*

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	5	3.54	Yes	7	3.07	Yes
		8	5	1.70	Yes	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	9	2.38	Yes	1	X	X
		8	1	X	X	1	X	X
		4	3	2.20	Yes	1	X	X
2	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
3	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
1 Variability Test	Cooperative	12	7	3.16	Yes	9	2.78	Yes
		12	8	2.75	Yes	6	3.71	Yes
		12	8	3.22	Yes	9	2.83	Yes
		12	7	3.80	Yes	1	X	X
		12	7	3.65	Yes	6	4.06	Yes
		12	8	2.93	Yes	6	4.94	Yes
		12	6	4.90	Yes	7	3.20	Yes
		12	5	5.85	Yes	6	5.09	Yes
1 Photo Test	Cooperative	12	5	6.03	Yes	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X

Table 13: *Miros (eTrue)—Old Image Database Timed Test Verification Mode*

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	1	7.49	Yes	4	4.68	Yes
		8	1	8.47	Yes	1	X	X
		4	1	4.16	Yes	1	4.19	Yes
	Indifferent	12	2	6.28	Yes	5	4.42	Yes
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
2	Cooperative	12	2	7.40	Yes	6	4.51	Yes
		8	1	X	X	1	X	X
		4	1	6.31	Yes	1	5.04	Yes
	Indifferent	12	6	4.47	Yes	4	4.99	Yes
		8	1	X	X	1	9.64	Yes
		4	1	X	X	1	X	X
3	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	8.81	Yes	1	7.23	Yes
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	4.23	Yes	1	X	X
1 Variability Test	Cooperative	12	4	4.44	Yes	4	4.88	Yes
		12	1	8.51	Yes	7	3.06	Yes
		12	2	6.47	Yes	8	3.26	Yes
		12	4	5.05	Yes	8	3.02	Yes
		12	7	3.79	Yes	8	3.46	Yes
		12	6	4.58	Yes	8	3.12	Yes
		12	1	8.99	Yes	6	4.78	Yes
		12	2	6.89	Yes	8	3.28	Yes
1 Photo Test	Cooperative	12	1	X	X	1	X	X
		8	2	5.24	Yes	4	3.58	Yes
		4	1	4.46	Yes	1	5.01	Yes

Table 14: *Visionics Corp.—Old Image Database Timed Test Verification Mode*

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	4.03	No	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	10.00	No
2	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	10.00	No
	Indifferent	12	1	X	X	1	X	X
		8	1	8.50	No	1	X	X
		4	1	8.87	No	1	10.00	No
3	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
1 Variability Test	Cooperative	12	3	5.14	No	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
1 Photo Test	Cooperative	12	1	8.11	No	1	X	X
		8	4	3.78	No	1	X	X
		4	4	2.74	No	1	X	X

Table 15: *Banque-Tec—Old Image Database Timed Test Identification Mode*

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	3	6.31	No	2	7.91	No
		8	4	3.44	No	1	8.02	No
		4	2	3.91	No	1	7.40	No
	Indifferent	12	6	4.54	No	3	7.12	No
		8	4	3.37	No	1	X	X
		4	2	4.29	No	1	4.69	No
2	Cooperative	12	5	5.44	No	1	8.35	No
		8	1	10.00	No	1	6.66	No
		4	1	5.03	No	1	4.87	No
	Indifferent	12	6	4.85	No	2	8.38	No
		8	1	5.93	No	1	6.49	No
		4	1	6.57	No	1	7.29	No
3	Cooperative	12	1	8.33	No	7	3.89	No
		8	1	8.05	No	1	9.43	No
		4	1	5.86	No	1	8.74	No
	Indifferent	12	6	4.52	No	6	4.75	No
		8	1	9.64	No	1	9.99	No
		4	1	8.57	No	1	6.25	No
1 Variability Test	Cooperative	12	1	9.43	No	6	4.55	No
		12	1	8.56	No	6	4.50	No
		12	3	6.54	No	5	5.01	No
		12	2	7.16	No	5	5.14	No
		12	1	7.19	No	5	5.34	No
		12	3	6.31	No	3	6.67	No
		12	5	4.12	No	1	10.00	No
		12	1	10.00	No	1	8.59	No
1 Photo Test	Cooperative	12	6	5.24	No	7	4.52	No
		8	2	4.79	No	3	4.87	No
		4	1	4.26	No	1	6.92	No

Table 16: *C-VIS—Old Image Database Timed Test Identification Mode*

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
2	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
3	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
1 Variability Test	Cooperative	12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
1 Photo Test	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X

Table 17: *Lau Technologies—Old Image Database Timed Test Identification Mode*

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
2	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
3	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
1 Variability Test	Cooperative	12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
1 Photo Test	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X

Table 18: *Miros (eTrue)—Old Image Database Timed Test Identification Mode*

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	1	10.00	Yes	1	9.44	Yes
		8	1	X	X	4	4.52	Yes
		4	1	X	X	1	X	X
2	Cooperative	12	1	X	X	1	X	X
		8	1	7.19	Yes	1	10.00	Yes
		4	1	X	X	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
3	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	Yes
		4	1	5.97	Yes	1	X	X
	Indifferent	12	1	10.00	Yes	1	X	X
		8	1	9.51	Yes	1	X	X
		4	2	5.66	Yes	1	4.82	Yes
1 Variability Test	Cooperative	12	1	8.18	Yes	1	X	X
		12	1	X	X	3	6.49	Yes
		12	1	X	X	3	6.87	Yes
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	3	7.14	Yes
		12	1	8.01	Yes	1	X	X
1 Photo Test	Cooperative	12	1	X	X	3	7.48	Yes
		8	1	X	X	1	X	X
		4	1	9.52	Yes	1	X	X

Table 19: *Visionics Corp.—Old Image Database Timed Test Identification Mode*

7.2.5 Enrollment Timed Test Results

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	12	X	X	12	X	X
		8	8	7.95	No	8	X	X
		4	4	1.47	Yes	4	1.29	Yes
	Indifferent	12	12	X	X	12	X	X
		8	8	3.23	Yes	8	3.02	Yes
		4	4	6.96	Yes	4	1.71	Yes
2	Cooperative	12	12	X	X	12	X	X
		8	8	7.86	No	8	X	X
		4	4	1.77	Yes	4	1.39	Yes
	Indifferent	12	12	10.00	No	12	X	X
		8	8	X	X	8	X	X
		4	4	2.10	Yes	4	1.72	Yes
3	Cooperative	12	12	X	X	12	X	X
		8	8	X	X	8	X	X
		4	4	1.89	Yes	4	7.07	No
	Indifferent	12	12	X	X	12	X	X
		8	8	10.00	Yes	8	7.83	No
		4	4	2.64	Yes	4	X	X
1 Variability Test	Cooperative	12	12	X	X	12	X	X
		12	12	X	X	12	X	X
		12	12	X	X	12	X	X
		12	12	X	X	12	8.12	No
		12	12	X	X	12	X	X
		12	12	X	X	12	X	X
		12	12	X	X	12	X	X
		12	12	X	X	12	X	X
1 Photo Test	Cooperative	12	12	X	X	12	X	X
		8	8	X	X	8	X	X
		4	4	7.47	No	4	7.92	No

Table 20: *Banque-Tec—Enrollment Timed Test Verification Mode*

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	12	10.00	Yes	12	3.94	Yes
		8	8	3.14	Yes	8	4.88	Yes
		4	4	5.92	Yes	4	8.42	No
	Indifferent	12	12	8.49	No	12	4.39	No
		8	8	3.86	Yes	8	5.73	Yes
		4	4	X	X	4	3.48	Yes
2	Cooperative	12	12	3.85	No	12	3.25	Yes
		8	8	3.06	Yes	8	6.07	Yes
		4	4	5.05	Yes	4	5.04	Yes
	Indifferent	12	12	2.85	Yes	12	3.82	Yes
		8	8	4.09	Yes	8	3.71	Yes
		4	4	5.45	Yes	4	5.69	Yes
3	Cooperative	12	12	3.81	No	12	3.25	Yes
		8	8	4.24	Yes	8	3.23	Yes
		4	4	4.01	Yes	4	4.03	Yes
	Indifferent	12	12	3.34	No	12	3.73	No
		8	8	8.19	Yes	8	8.59	Yes
		4	4	10.00	Yes	4	4.04	Yes
1 Variability Test	Cooperative	12	12	4.01	No	12	2.76	Yes
		12	12	5.00	Yes	12	5.30	Yes
		12	12	3.62	No	12	3.55	Yes
		12	12	4.04	No	12	3.50	Yes
		12	12	4.79	Yes	12	3.89	No
		12	12	2.93	Yes	12	3.86	Yes
		12	12	3.92	Yes	12	3.31	Yes
		12	12	3.48	Yes	12	2.83	Yes
1 Photo Test	Cooperative	12	12	4.88	No	12	2.85	No
		8	8	X	X	8	3.79	No
		4	4	6.63	No	4	5.69	No

Table 21: *C-VIS—Enrollment Timed Test Verification Mode*

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	12	1.78	Yes	12	1.50	Yes
		8	8	2.07	Yes	8	1.05	Yes
		4	4	1.25	Yes	4	0.90	Yes
	Indifferent	12	12	1.44	Yes	12	1.67	Yes
		8	8	1.29	Yes	8	1.37	Yes
		4	4	2.46	Yes	4	X	X
2	Cooperative	12	12	1.56	Yes	12	1.46	Yes
		8	8	1.04	Yes	8	0.93	Yes
		4	4	1.54	Yes	4	2.24	Yes
	Indifferent	12	12	1.50	Yes	12	7.68	Yes
		8	8	1.80	Yes	8	1.69	Yes
		4	4	1.85	Yes	4	1.53	Yes
3	Cooperative	12	12	3.47	Yes	12	X	X
		8	8	1.11	Yes	8	4.63	Yes
		4	4	1.18	Yes	4	1.30	Yes
	Indifferent	12	12	2.71	Yes	12	2.70	Yes
		8	8	1.19	Yes	8	1.22	Yes
		4	4	1.47	Yes	4	0.96	Yes
1 Variability Test	Cooperative	12	12	8.14	Yes	12	1.14	Yes
		12	12	X	X	12	2.76	Yes
		12	12	1.32	Yes	12	0.89	Yes
		12	12	1.19	Yes	12	1.95	Yes
		12	12	1.52	Yes	12	1.42	Yes
		12	12	2.54	Yes	12	2.08	Yes
		12	12	0.77	Yes	12	1.28	Yes
		12	12	0.96	Yes	12	1.90	Yes
1 Photo Test	Cooperative	12	12	X	X	12	X	X
		8	8	X	X	8	X	X
		4	4	X	X	4	X	X

Table 22: *Lau Technologies—Enrollment Timed Test Verification Mode*

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	12	3.01	Yes	12	X	X
		8	8	1.57	Yes	8	1.65	Yes
		4	4	1.62	Yes	4	X	X
	Indifferent	12	12	2.10	Yes	12	3.05	Yes
		8	8	2.16	Yes	8	1.62	Yes
		4	4	3.69	Yes	4	X	X
2	Cooperative	12	12	X	X	12	10.00	Yes
		8	8	1.98	Yes	8	1.49	Yes
		4	4	8.35	Yes	4	X	X
	Indifferent	12	12	2.60	Yes	12	9.48	Yes
		8	8	2.20	Yes	8	3.22	Yes
		4	4	9.68	Yes	4	X	X
3	Cooperative	12	12	2.48	Yes	12	X	X
		8	8	1.57	Yes	8	1.33	Yes
		4	4	X	X	4	X	X
	Indifferent	12	12	3.09	Yes	12	X	X
		8	8	2.48	Yes	8	1.50	Yes
		4	4	X	X	4	X	X
1 Variability Test	Cooperative	12	12	2.37	Yes	12	10.00	Yes
		12	12	2.19	Yes	12	5.73	Yes
		12	12	1.49	Yes	12	1.72	Yes
		12	12	1.73	Yes	12	2.15	Yes
		12	12	1.82	Yes	12	2.67	Yes
		12	12	1.86	Yes	12	2.19	Yes
		12	12	1.61	Yes	12	2.21	Yes
		12	12	1.55	Yes	12	2.60	Yes
1 Photo Test	Cooperative	12	12	X	X	12	7.23	Yes
		8	8	6.45	Yes	8	X	X
		4	4	2.25	Yes	4	2.33	Yes

Table 23: *Miros (eTrue)—Enrollment Timed Test Verification Mode*

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	12	6.59	Yes	12	X	X
		8	8	3.19	Yes	8	4.62	Yes
		4	4	2.91	Yes	4	3.89	Yes
	Indifferent	12	12	7.62	Yes	12	X	X
		8	8	2.54	Yes	8	4.83	Yes
		4	4	7.51	Yes	4	9.48	Yes
2	Cooperative	12	12	2.94	Yes	12	3.50	Yes
		8	8	3.02	Yes	8	2.58	Yes
		4	4	2.84	Yes	4	3.04	Yes
	Indifferent	12	12	2.87	Yes	12	3.39	Yes
		8	8	2.63	Yes	8	2.85	Yes
		4	4	2.99	Yes	4	2.78	Yes
3	Cooperative	12	12	3.27	Yes	12	3.54	Yes
		8	8	2.89	Yes	8	2.72	Yes
		4	4	3.01	Yes	4	2.90	Yes
	Indifferent	12	12	3.85	Yes	12	2.63	Yes
		8	8	2.63	Yes	8	2.76	Yes
		4	4	2.88	Yes	4	3.08	Yes
1 Variability Test	Cooperative	12	12	3.35	Yes	12	X	X
		12	12	2.48	Yes	12	X	X
		12	12	3.93	Yes	12	3.39	Yes
		12	12	3.01	Yes	12	6.67	Yes
		12	12	X	X	12	8.07	Yes
		12	12	4.24	Yes	12	X	X
		12	12	6.54	Yes	12	X	X
		12	12	2.72	Yes	12	9.37	Yes
1 Photo Test	Cooperative	12	12	X	X	12	X	X
		8	8	X	X	8	X	X
		4	4	X	X	4	X	X

Table 24: *Visionics Corp.—Enrollment Timed Test Verification Mode*

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	12	X	X	12	X	X
		8	8	X	X	8	X	X
		4	4	2.62	Yes	4	2.37	Yes
	Indifferent	12	12	X	X	12	X	X
		8	8	3.08	Yes	8	3.00	Yes
		4	4	1.70	Yes	4	2.53	Yes
2	Cooperative	12	12	8.63	No	12	X	X
		8	8	X	X	8	X	X
		4	4	2.09	Yes	4	1.58	Yes
	Indifferent	12	12	7.32	No	12	X	X
		8	8	X	X	8	X	X
		4	4	2.57	Yes	4	2.64	Yes
3	Cooperative	12	12	X	X	12	X	X
		8	8	X	X	8	X	X
		4	4	3.61	Yes	4	2.91	Yes
	Indifferent	12	12	X	X	12	X	X
		8	8	X	X	8	X	X
		4	4	2.48	Yes	4	10.00	No
1 Variability Test	Cooperative	12	12	X	X	12	X	X
		12	12	X	X	12	X	X
		12	12	X	X	12	X	X
		12	12	X	X	12	X	X
		12	12	X	X	12	X	X
		12	12	X	X	12	X	X
		12	12	X	X	12	X	X
		12	12	X	X	12	X	X
1 Photo Test	Cooperative	12	12	X	X	12	8.19	No
		8	8	X	X	8	7.60	No
		4	4	X	X	4	7.58	No

Table 25: *Banque-Tec—Enrollment Timed Test Identification Mode*

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	12	3.80	Yes	12	3.69	Yes
		8	8	4.42	Yes	8	4.79	Yes
		4	4	6.51	Yes	4	6.28	No
	Indifferent	12	12	7.81	Yes	12	3.71	Yes
		8	8	5.50	Yes	8	4.11	Yes
		4	4	10.00	Yes	4	4.94	Yes
2	Cooperative	12	12	3.67	Yes	12	3.78	Yes
		8	8	3.93	Yes	8	5.43	Yes
		4	4	3.72	Yes	4	6.80	Yes
	Indifferent	12	12	3.98	Yes	12	4.85	Yes
		8	8	4.02	Yes	8	4.69	Yes
		4	4	5.06	Yes	4	6.20	Yes
3	Cooperative	12	12	4.80	Yes	12	4.88	Yes
		8	8	6.33	Yes	8	4.38	Yes
		4	4	X	X	4	10.00	Yes
	Indifferent	12	12	6.49	Yes	12	5.04	Yes
		8	8	9.03	Yes	8	5.43	No
		4	4	10.00	Yes	4	8.72	No
1 Variability Test	Cooperative	12	12	4.41	Yes	12	3.76	Yes
		12	12	4.45	Yes	12	5.36	Yes
		12	12	5.06	Yes	12	3.71	Yes
		12	12	3.78	Yes	12	4.32	Yes
		12	12	4.33	Yes	12	3.78	Yes
		12	12	6.56	Yes	12	4.76	Yes
		12	12	10.00	Yes	12	3.77	Yes
		12	12	4.20	Yes	12	4.00	Yes
1 Photo Test	Cooperative	12	12	5.56	No	12	5.07	No
		8	8	6.24	No	8	6.61	No
		4	4	8.50	No	4	7.22	No

Table 26: *C-VIS—Enrollment Timed Test Identification Mode*

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	12	3.15	Yes	12	2.90	Yes
		8	8	2.10	Yes	8	1.67	Yes
		4	4	2.43	Yes	4	1.32	Yes
	Indifferent	12	12	2.17	Yes	12	2.21	Yes
		8	8	1.96	Yes	8	5.44	Yes
		4	4	6.47	Yes	4	X	X
2	Cooperative	12	12	2.27	Yes	12	1.47	Yes
		8	8	1.60	Yes	8	1.52	Yes
		4	4	1.81	Yes	4	1.23	Yes
	Indifferent	12	12	2.33	Yes	12	X	X
		8	8	2.23	Yes	8	1.38	Yes
		4	4	2.59	Yes	4	1.29	Yes
3	Cooperative	12	12	2.09	Yes	12	2.30	Yes
		8	8	1.57	Yes	8	3.42	Yes
		4	4	2.29	Yes	4	2.25	Yes
	Indifferent	12	12	1.78	Yes	12	4.54	Yes
		8	8	2.02	Yes	8	X	X
		4	4	2.40	Yes	4	2.39	Yes
1 Variability Test	Cooperative	12	12	1.97	Yes	12	3.04	Yes
		12	12	2.24	Yes	12	4.50	Yes
		12	12	2.25	Yes	12	2.68	Yes
		12	12	3.95	Yes	12	4.88	Yes
		12	12	2.57	Yes	12	3.32	Yes
		12	12	2.83	Yes	12	3.06	Yes
		12	12	2.73	Yes	12	3.15	Yes
		12	12	1.87	Yes	12	3.17	Yes
1 Photo Test	Cooperative	12	12	X	X	12	X	X
		8	8	X	X	8	X	X
		4	4	X	X	4	X	X

Table 27: *Lau Technologies—Enrollment Timed Test Identification Mode*

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	12	5.06	Yes	12	6.12	Yes
		8	8	6.50	Yes	8	2.97	Yes
		4	4	3.45	Yes	4	5.21	Yes
	Indifferent	12	12	4.67	Yes	12	X	X
		8	8	9.96	Yes	8	4.70	Yes
		4	4	6.72	Yes	4	4.73	Yes
2	Cooperative	12	12	X	X	12	4.74	Yes
		8	8	3.41	Yes	8	2.63	Yes
		4	4	5.43	Yes	4	8.89	Yes
	Indifferent	12	12	4.59	Yes	12	X	X
		8	8	7.01	Yes	8	5.58	Yes
		4	4	4.68	Yes	4	X	X
3	Cooperative	12	12	5.66	Yes	12	X	X
		8	8	3.65	Yes	8	6.69	Yes
		4	4	6.43	Yes	4	X	X
	Indifferent	12	12	4.48	Yes	12	X	X
		8	8	3.49	Yes	8	3.39	Yes
		4	4	X	X	4	X	X
1 Variability Test	Cooperative	12	12	5.29	Yes	12	3.62	Yes
		12	12	6.67	Yes	12	3.14	Yes
		12	12	3.75	Yes	12	7.50	Yes
		12	12	4.63	Yes	12	X	X
		12	12	4.76	Yes	12	X	X
		12	12	7.30	Yes	12	4.13	Yes
		12	12	3.89	Yes	12	5.74	Yes
		12	12	6.39	Yes	12	7.96	Yes
1 Photo Test	Cooperative	12	12	X	X	12	X	X
		8	8	X	X	8	X	X
		4	4	X	X	4	X	X

Table 28: *Miros (eTrue)—Enrollment Timed Test Identification Mode*

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	12	X	X	12	X	X
		8	8	8.09	Yes	8	8.74	Yes
		4	4	X	X	4	8.28	Yes
	Indifferent	12	12	X	X	12	X	X
		8	8	6.59	Yes	8	5.66	Yes
		4	4	8.79	Yes	4	X	X
2	Cooperative	12	12	X	X	12	9.04	Yes
		8	8	8.88	Yes	8	9.23	Yes
		4	4	10.00	Yes	4	9.66	Yes
	Indifferent	12	12	8.64	Yes	12	8.52	Yes
		8	8	9.32	Yes	8	7.67	Yes
		4	4	8.20	Yes	4	X	X
3	Cooperative	12	12	X	X	12	X	X
		8	8	8.38	Yes	8	8.25	Yes
		4	4	8.12	Yes	4	8.87	Yes
	Indifferent	12	12	8.36	Yes	12	8.72	Yes
		8	8	9.19	Yes	8	7.54	Yes
		4	4	9.77	Yes	4	9.80	Yes
1 Variability Test	Cooperative	12	12	X	X	12	X	X
		12	12	X	X	12	X	X
		12	12	8.60	Yes	12	X	X
		12	12	9.57	Yes	12	X	X
		12	12	X	X	12	X	X
		12	12	X	X	12	X	X
		12	12	X	X	12	8.70	Yes
		12	12	X	X	12	9.81	Yes
1 Photo Test	Cooperative	12	12	X	X	12	X	X
		8	8	X	X	8	X	X
		4	4	X	X	4	X	X

Table 29: *Visionics Corp.—Enrollment Timed Test Identification Mode*

8 Lessons Learned for Future Evaluations

8.1 Vendor Comments

Each vendor was asked to suggest any improvements they would like to see in future evaluations. A summary of those suggestions follows.

Recognition Performance Test:

- Provide more than 18 images in the sample set to demonstrate more variations.
- Report face-finding coordinates in the similarity files to allow a separate evaluation of face finding and matching.
- Use inexpensive hard drives to store the similarity files rather than expensive Jaz disks.

Product Usability Test:

- Use video instead of live subjects to ensure consistency.
- Add tests for imposters.
- Add tests with lighting at side and bottom of subjects to fully test the effects of lighting variation.
- Test with multiple subjects in field of view.
- Test each product according to intended application.

8.2 Sponsor Comments

The sponsors of the FRVT 2000 spent a considerable amount of time planning these evaluations and tried to counter any potential problems before they arose. Because of the magnitude of these evaluations and the fact they were being performed on commercial systems, the sponsors also understood that unforeseen issues would arise. It is as important to document the background work and any obstacles that were encountered as it is to document the results of the vendor evaluations. Most of these items have been covered in previous sections of this report. Some did not have a natural fit with the other subject matter and have been placed in the following subsections.

8.2.1 Lessons Learned Before the Evaluation Dates

Because of the lessons learned from previous scenario evaluations (described in Section 3.4), the sponsors provided a detailed overview of the format of the FRVT 2000 evaluations in the overview page on the FRVT 2000 web site. It seemed likely that the vendors would propose modifications to the evaluation protocol because the FERET program participants did also. This issue was successfully settled at the start by addressing this in the FAQ section of the web site as shown below:

25. Can my company propose changes to the planned tests?

Absolutely. We are always looking for new ideas on how to compare one system to another. The sponsors, however, spent considerable time developing the test plan for the Facial Recognition Vendor Test 2000 and decided that the method given on this web site is how the tests will be performed. It would be unfair to other test participants to change the tests at this point. We will gladly hold on to all proposed changes and study them if we should do another series of tests.

This approach proved to be effective as only one vendor voiced objections regarding the evaluation methodology. This vendor eventually withdrew from the evaluation. The letter requesting to withdraw from the evaluations stated that the reason was their disapproval of the evaluation method used in FRVT 2000. When the sponsors received this letter via e-mail, they sent a reply granting the vendor's request and also described the validity of the evaluation method.

The sponsors did not find out until much later that the vendor also sent copies of the withdrawal request e-mail to all of the other vendors participating in the FRVT 2000. The message was sent separately to the other vendors, so the sponsors therefore, did not copy any of the other vendors on their reply letter to this vendor.

When seen from the viewpoint of the other participants, one vendor had questioned the validity of the evaluation method in an apparently open forum without the evaluation sponsors responding whatsoever. In hindsight, the sponsors feel that this may have had a negative effect as two other vendors subsequently withdrew within the next 36 hours. Fortunately, one of these vendors requested to rejoin the evaluation the following week.

The lesson learned from this chain of events is that all discussions with anyone outside those running the evaluation should be completely open to the public. The sponsors had worked to ensure that the participating vendors had a level playing field via the Q&A restrictions but, in this case, a further degree of restrictions on discussion would have been beneficial.

8.2.2 Product Usability Test

The sponsors did not expect the disparity in performance found when comparing the Old Image Database Timed Test and the Enrollment Timed Test. Although it was expected that the systems would perform better in the Enrollment Timed Test, the performance in the Old Image Database Timed Test was worse than expected. In future evaluations, it would be beneficial to add a third timed test to allow the vendors to enroll the subjects as they desire but in a different room with different lighting conditions than where the tests were performed. It is expected that this test would give results somewhere between the results of the Old Image Database Timed Test and the Enrollment Timed Test.

During the photo test, an 8" x 10" glossy color photograph was used that showed a bright spot from the reflections of the overhead lights. This was compounded by the fact that it was not mounted on a rigid structure. If the photo was bent, the glare was more severe. The subject holding the photo made an active effort to minimize this effect by keeping it parallel to the plane of the camera and pulling outward on the edges to keep it straight. We recommend using a matte-finish photo mounted on rigid support for future evaluations.

9 Summary

The Facial Recognition Vendor Test 2000 has been a worthwhile endeavor. It will help numerous readers evaluate facial recognition systems for their own uses. The sponsors have learned a great deal about the status of commercially available facial recognition systems, evaluation methodologies and vendor business practices. The sponsors hope that this knowledge has been conveyed to the biometrics community through this report.

The FRVT 2000 evaluations were not designed, and this report was not written, to be a buyer's guide for facial recognition. No one will be able to open this report to a specific page to determine which facial recognition system is best because there is not one system for all applications. The only

way to determine the best facial recognition system for any application is to follow the three-step evaluation methodology described in this report and analyze the data as it pertains to each individual application. It is possible that some of the experiments performed in the Recognition Performance and Product Usability portions of this evaluation have no relation to a particular application and should be ignored.

9.1 Compression Experiments

The compression experiments show that compression of facial images does not necessarily adversely affect performance. Results presented in figure 7 show that performance increased slightly for 10:1 and 20:1 compression rates versus uncompressed probe images. It is not until a compression ratio of 40:1 that the performance rate drops below that of the uncompressed probes. Because the results are aggregated and only consider JPEG compression, we recommend that additional studies on the effect of compression on face recognition systems be conducted.

9.2 Pose Experiments

The pose experiments show that performance is stable when the angle between a frontal gallery image and a probe is less than 25 degrees and that performance dramatically falls off when the angle is greater than 40 degrees.

9.3 Temporal Experiments

For the FERET temporal probe sets, the FRVT 2000 performance for the duplicate I (T1) and duplicate II (T2) probes have almost the same top rank score. (The duplicate I probes are probes taken on different days or under different conditions than the gallery images; the duplicate II probes and gallery images were taken at least 18 months apart.) In the FERET 1996 evaluation, the algorithms evaluated performed better on the duplicate I probe set. In the FERET evaluations, there was approximately a seven percentage point difference in performance between duplicate I and II probes for the best partially automatic algorithm.

The T3, T4 and T5 experiments use the same probe set and vary the type of images in the gallery. The time between the collection of the gallery and probe images was at least one year. The T3, T4 and T5 experiments are similar to the FERET duplicate II probe set (T2 experiment) because there was at least one year between the time the gallery and probe images were acquired. The gallery in T3 consisted of images taken with best-practice mugshot lighting, the T4 gallery contained FERET-style images and the T5 gallery's images were taken with overhead lighting. Based on the top match score, the hardest experiment was T5; the easiest was T3. The verification scores do not produce such a ranking of the experiments. The top identification scores were 0.55 for T3, 0.55 for T4 and .35 for T5, which are lower than the best T2 top match score of 0.65. The temporal results show that recognizing faces from images taken more than a year apart remains an active area of research.

9.4 Distance Experiments

The distance experiments across all algorithms and the three sets of distance experiments show that performance decreased as distance between the person and camera increased. There were three sets of distance experiments: experiments D1-D3 (indoor digital gallery images, indoor video probes 2, 3 and 5 meters from the camera), D4 and D5 (indoor video gallery images, indoor video probes 3 and 5 meters from the camera) and, D6 and D7 (outdoor video gallery images, outdoor video probes 3 and 5 meters from the camera).

9.5 Expression Experiments

For the identification performance in the expression experiment, all three algorithms performed better on the E1 case; whereas during verification, all three algorithms achieved their best performance on the E2 case. The difference in identification performance between the E1 and E2 cases for the top match score ranged from three to five percentage points, and zero to two percentage points for the verification equal error rate. This shows that for the FRVT 2000, identification is more sensitive to changes in expression than verification.

9.6 Illumination Experiments

In the illumination experiment, the I3 case was the most difficult and I2 was the least difficult. Illumination experiments I1 and I3 used the same gallery of digital mugshots taken indoors, the I1 probe set had indoor digital images with overhead lighting, and the I3 probe set's images were taken outdoors. The I1 experiment's performance was significantly better than the I3 experiment, which shows that an area of future investigation is handling lighting changes that occur when one image is taken indoors and the other is taken outdoors.

9.7 Media Experiments

For Lau Technologies and Visionics Corp., switching between media did not significantly affect performance. For case M1, the gallery consisted of 35mm images and the probe set consisted of digital images. For the M2 case, the gallery contained digital images and the probe set 35mm images.

9.8 Resolution Experiments

In this experiment, the R2 performance values were better than the R1 scores except for the verification performance of C-VIS. (The inter-pupil distance for R1 was 60 pixels and 45 pixels for R2.) All systems had their worst performance on the R4 case (inter-pupil distance of 15).

9.9 Overall Conclusions for the Recognition Performance Test

The FERET evaluations identified temporal and pose variations as two key areas for future research in face recognition. The FRVT 2000 shows that progress has been made in temporal changes, but developing algorithms that can handle temporal variations is still a necessary research area. In addition, developing algorithms that can compensate for pose variations, and illumination and distance changes were noted as other areas for future research.

The FRVT 2000 experiments on compression confirm the findings of Moon and Phillips that moderate levels of compression do not adversely affect performance. The resolution experiments find that moderately decreasing the resolution can slightly improve performance. In most cases, compression and reducing resolution are lowpass filters. Both results suggest that low-pass filtering probes could increase performance.

9.10 Product Usability Test

In the Product Usability Tests, all vendors performed considerably better in the Enrollment Timed Tests than in the Old Image Database Timed Tests. There are two main differences between the two tests. The first is that the subjects are walking towards the camera in the Old Image Database Timed Test and are stationary for the Enrollment Timed Tests. Results from the Recognition Performance Test show us that performance actually increases as the subjects get closer to the camera, so this would not cause the degradation in performance seen in the Old Image Database Timed Test.

The second difference between the two Product Usability Tests is the enrollment method of gallery images. In the Old Image Database Test, the gallery images were provided to the vendors before testing began. These images were taken with different camera systems and in a different location than where the testing occurred. In the Enrollment Timed Test, the gallery images were enrolled using the vendor system and in the same room where testing took place. By default, this difference in enrollment procedures is the cause of the change in performance by the systems in the Product Usability Tests. This shows that potential users of facial recognition technology should enroll subjects using images gathered by the facial recognition system at the installation location if at all possible. These results also show facial recognition vendors that this is an area for additional research.

In all cases, there was very little difference in performance between cooperative and simulated indifferent results. The lack of a difference is mainly because of the pose angles introduced by the simulated indifferent behavior. The initial pose angle varied between 17 and 24 degrees, depending on the start distance, and decreased as the subject began simulating indifferent behavior. These results are in agreement with the pose experiments in the Recognition Performance Test and show that facial recognition systems will not show significant changes in performance if a subject is cooperative versus indifferent as long as the indifferent subject is facing toward the camera.

Adding moderate, non-varying backlighting generally introduced a small degree of difficulty for the facial recognition vendors, but in most cases it was negligible. Further experimentation with higher intensity backlighting, lighting from various angles and varying intensity are necessary to fully understand the impact of lighting in this scenario.

In all cases, the facial recognition systems were quicker and more accurate when performing verification experiments than in identification experiments. The gallery size for identification experiments in the Product Usability Test was 165, which is a fairly small number. It is anticipated that performance disparity will increase as the identification gallery increases, but further tests are required to know for sure.

Two of the five companies correctly returned no score for the photo tests in the Enrollment Timed Test. This evaluation was a very quick look at the "liveness" issue that is important for any form of access control using biometrics, but it may not be an issue for other applications. Additional research on this issue should be carried out for the three systems that attempted to identify the individual and on the two that correctly returned no score to determine their consistency.

The sponsors are already using the knowledge gained, the databases and scoring algorithms from FRVT 2000 for numerous development, evaluation, and demonstration programs. The sponsors look forward to learning, during the next several months, how others are using this report and want to thank the community for the privilege of providing this service to them.

Appendix O – Participant's Comments on the FRVT 2000 Evaluation Report